# Schriftliche Ausarbeitungen für das Seminar Internetdienste (Sommersemester 2003)

**Universität Ulm**
**Abteilung Angewandte Informationsverarbeitung**

**und**

**DaimlerChrysler Forschung & Technologie, Ulm**
**Abteilung Information Mining (Ric/AM)**

Ulm, im August 2003

# Schriftliche Ausarbeitungen für das Seminar Internetdienste (Sommersemester 2003)

Das Seminar wurde in enger Zusammenarbeit zwischen der Abteilung Angewandte Informationsverarbeitung (Prof. Dr. Franz Schweiggert) und dem DaimlerChrysler Forschungsinstitut Ulm (Thilo Maier, Dr. Ingo Melzer, Dr. Matthias Grabert und Prof. Dr. Reza Nakhaeizadeh) durchgeführt. Im folgenden finden sich die einzelnen schriftlichen Ausarbeitungen. Da es sich um eine Sammlung von Einzelwerken handelt, liegt leider *keine* Gesamtseitennummerierung vor. Die einzelnen Arbeiten befinden sich in der nachstehend aufgeführten Reihenfolge:

**1. Themenblock (Internet & Gesellschaft)**

| Thema | Vortragender |
|---|---|
| Netiquette | J. Czeschner |
| Standards | G. Gröner |
| Rechtsfragen (1) | M. Entenmann |
| Rechtsfragen (2) | A. Kulas |

**2. Themenblock (Technologische Grundlagenthemen)**

| Thema | Vortragender |
|---|---|
| Firewall | F. Weidlich |
| Anonymisierung | J. Bauer |
| Grid-Computing | N. Heidenbluth |

**3. Themenblock (Technologische AddOns & Data Mining)**

| Thema | Vortragender |
|---|---|
| Semantic Web | S. Beiler |
| Web Services | M. Kirchmayr |
| Web-Reporting | M. Seitz |
| PMML | M. Prechtel |

**4. Themenblock (Internet & Commerce)**

| Thema | Vortragender |
|---|---|
| E-Marketing | D. Niessner |
| E-Commerce | S. Reichardt |

# Die Netiquette, erstellt im Rahmen eines Seminars über Internet Dienste an der Univeristät Ulm.

Jürgen Czeschner

Version vom 11. August 2003

# Inhaltsverzeichnis

**1 Einleitung** **3**

**2 Die klassische Netiquette, RFC 1855** **4**
2.1 Elektronische Post oder one-to-one communication . . . . . 4
2.2 Telnet und Anonymes Fileübertragungsprotokoll (ftp) oder service communications . . . . . 4
2.3 Elektonische Kommunikation oder one-to-many-communication . . . . . 5
2.3.1 Vergessen Sie niemals, dass auf der anderen Seite ein Mensch sitzt! . . . . . 5
2.3.2 Erst lesen, dann denken. Noch einmal lesen, noch einmal denken. Und dann erst posten! . . . . . 5
2.3.3 Teilen Sie etwas Neues mit! . . . . . 5
2.3.4 Ihre Artikel sprechen für Sie - Seien Sie stolz auf sie! . . . 5
2.3.5 Nehmen Sie sich Zeit, wenn Sie einen Artikel schreiben! . 5
2.3.6 Vernachlässigen Sie nicht die Aufmachung Ihres Artikels! 6
2.3.7 Achten Sie auf die "Subject:"-Zeile! . . . . . 6
2.3.8 Denken Sie an die Leserschaft! . . . . . 6
2.3.9 Vorsicht mit Humor, Ironie und Sarkasmus! . . . . . 6
2.3.10 Kürzen Sie zitierten Text auf das notwendige Minimum! . 6
2.3.11 Benutzen Sie E-Mail! . . . . . 7
2.3.12 Geben Sie eine Sammlung Ihrer Erkenntnisse an das Netz weiter! . . . . . 7
2.3.13 Achten Sie auf die gesetzlichen Regelungen! . . . . . 7
2.3.14 Benutzen Sie Ihren wirklichen Namen, kein Pseudonym! . 7
2.3.15 Vorsicht mit Kommerziellem! . . . . . 7
2.3.16 Vorsicht mit Binaries und Multipart-Artikeln! . . . . . 7
2.3.17 "Du" oder "Sie". . . . . 7
2.4 Die Netiquette der Webseitenautoren . . . . . 8

**3 Probleme, Diskussionen und Aktualisierungen der (.de*)-Netiquette** **8**

**4 Meine Version der (.de*)-Netiquette für deutschsprachige Newsgroups** **9**

**5 Ausblick** **13**

**6 Literaturverzeichnis** **13**

# 1 Einleitung

Die Netiquette (stammt von "Net-Etikette") ist das Regelwerk im Bereich der Datenkommunikation und enthält Grundregeln zum Umgang mit anderen Netzteilnehmern. Sie verbietet unter anderem persönliche Beleidigung und grobe Verletzung religiöser, weltanschaulicher oder auch ethischer Empfindungen anderer Netzteilnehmer, kommerzielle oder politische Werbung, rassistische und faschistische Äußerungen, Aufforderungen zu Gewalttaten und kriminellen Delikten. Kurz gesagt spiegelt die Netiquette den gesunden Menschenverstand wieder, also der verantwortungsvolle und sinnvolle Umgang mit Datenkommunikationsmitteln.

Man darf die Netiquette aber nicht als Gesetz oder strenge Vorschrift interpretieren, sondern eher als Hilfestellung, Nachschlagewerk und Richtlinie.

Es gibt viele verschiedene Versionen der Netiquette, wie die (.de*)-Netiquette, welche speziell für die deutschsprachigen Newsgroups ausgelegt ist. Eine weitere Netiquette im Umgang mit Newsgroups ist die (schule.*)-Netiquette [1], in der einige Erweiterungen der allgemeinen (.de*)-Netiquette, hinsichtlich der Nutzung des Schulnetzes, eingebracht wurden. Das meistverbreitestete und international anerkannte Regelwerk ist die RFC [2]1855, welche nicht nur Regeln im Umgang mit Newsgroups beinhaltet, sondern auch Richtlinien im Bereich der elektronischen Post (one-to-one communication), Telnet und ftp (information services) und der elektronischen Kommunikation (one-to-many communication). Ebenso werden dort die 10 Gebote der Computerethik aufgestellt.

**Die Zehn Gebote der Computerethik:**

1. Du sollst nicht Deinen Computer benutzen, um anderen Schaden zuzufügen.
2. Du sollst nicht anderer Leute Arbeit am Computer behindern.
3. Du sollst nicht in anderer Leute Files stöbern.
4. Du sollst nicht den Computer zum Stehlen benutzen.
5. Du sollst nicht den Computer benutzen, um falsches Zeugnis abzulegen.
6. Du sollst nicht Software benutzen oder kopieren, für die Du nicht gezahlt hast.
7. Du sollst nicht anderer Leute Ressourcen ohne deren Erlaubnis verwenden.
8. Du sollst nicht anderer Leute geistig Werk als Deines ausgeben.
9. Du sollst über die sozialen Konsequenzen Deiner Programme nachdenken.
10. Du sollst den Computer so benutzen, dass Du Verantwortung und Respekt zeigst.

[1] nachzulesen unter ftp.rus.uni-stuttgart.de:/pub/doc/faq/de.newusers/... oder in der Newsgroup schule.allgemein

[2] Request for Comment, (internationaler Standard)

# 2 Die klassische Netiquette, RFC 1855

Erstellt von Arlene H. Rinaldi Academic/Institutional Support Services Florida Atlantic University Juli, 1994

## 2.1 Elektronische Post oder one-to-one communication

Jeder Nutzer ist für Inhalt und Aufrechterhaltung (Wartung) seiner Mailbox verantwortlich, indem man täglich seine E-mails überprüft und gegebenfalls unerwünschte Nachrichten sofort entfernt. Generell sollten so wenige Mails wie möglich in der Mailbox liegen, um freien Speicherplatz zu schaffen, und deshalb sollten Mails öfter mal auf die private Festplatte kopieren werden. Zum Verfassen und Verwalten von Mails gilt der Grundsatz: Schicke nichts und hebe nichts auf, was Dich stören würde, wenn es Dritte lesen würden. Denke auch immer daran, dass das Internet nicht sicher ist, und deshalb auch andere Deine Mails lesen können. Kettenbriefe sind zu unterlassen, ebenso wie provozierende und beleidigende Nachrichten. Eine bis zu 4-zeilige Signatur, als Visitenkarte, wird gerne gesehen, damit man genauere Informationen über seinen "Gesprächspartner" erfährt. Mails mit einer Länge von über 100 Zeilen sollten in der Subject-Zeile mit "long" gekennzeichnet werden, und Anhänge sollten die Größe von 50 Kilobyte nicht überschreiten. Ebenso sollte man eine angemessene Beantwortungs-Zeit respektieren, welche bis zu einer Woche betragen kann.

## 2.2 Telnet und Anonymes Fileübertragungsprotokoll (ftp) oder service communications

Bei auftretenden Problemen sollte nicht sofort der Provider kontaktiert werden, sondern erst die eigenen Ressourcen überprüfen werden. Die Typbeschreibungen beim Download müssen nicht immer mit den gängigen Typen übereinstimmen. Ebenfalls ist auf mögliche Dokumentationsfiles, wie readme, zu achten. Wie überall im Umgang mit dem Internet ist man Sicherheitsmängeln ausgesetzt, deshalb sollte man alle Downloads auf Viren überprüfen und bei Ihrer ausführen kritisch die Vorgänge verfolgen. Weiterhin sollte auch an andere Benutzer gedacht werden. Auch sollte man sich nur so lange in einem System aufhalten, wie unbedingt notwendig und es anschließend verlassen. Größere Files (größer als 1 MB) sollten, wenn möglich spät abends übertragen werden. Ganz wichtig sind auch hier auf gesetzliche Bestimmungen, wie Urheberrechte und Lizenzbestimmungen zu achten. Sollten Zweifel bestehen, dann lieber nicht kopieren.

## 2.3 Elektonische Kommunikation oder one-to-many-communication

In diesem Abschnitt geht es um die Verhaltensregeln in Newsgroups und Mailinglists. Die RFC 1855 überschneidet sich in diesem Bereich mit der deutschsprachigen Netiquette für Newsgroups, die (.de*)-Netiquette, welche in diesem Abschnitt die internationale Netiquette ersetzt.

Die (.de*)-Netiquette besteht aus 17, hierarchisch angeordneten Regeln, auf welche im Folgenden etwas genauer eingegangen wird.

### 2.3.1 Vergessen Sie niemals, dass auf der anderen Seite ein Mensch sitzt!

Die Nachrichten und Artikel werden nicht von einer Maschine ausgewertet oder gelesen, sondern ebenfalls von Menschen. Was implizit auf die Beachtung von Höflichkeitsregeln hinweisen soll.

Deshalb eine einfache Faustregel: Nichts schreiben, was man dem Adressaten nicht auch vor anderen Leuten ins Gesicht sagen würde.

### 2.3.2 Erst lesen, dann denken. Noch einmal lesen, noch einmal denken. Und dann erst posten!

Sehr wichtig bei der schriftlichen Kommunikation, um die Gefahr von Missverständnissen vorzubeugen. Will der Autor wirklich das ausdrücken, was ich denke, oder wurde etwas ironisch oder fälschlich dargestellt, ohne gekennzeichnet zu sein?

### 2.3.3 Teilen Sie etwas Neues mit!

Man sollte einen klaren Standpunkt vertreten und sich verständlich und präzise ausdrücken.

### 2.3.4 Ihre Artikel sprechen für Sie - Seien Sie stolz auf sie!

Die Meinungsbildung der einzelnen User untereinander findet über die einzelnen Artikel statt. Man sollte also auf Stil, Form und gute Wortwahl achten, um als ernstzunehmende Person geachtet zu werden.

### 2.3.5 Nehmen Sie sich Zeit, wenn Sie einen Artikel schreiben!

Vor dem Abschicken eines Artikels sollte man diesen mindestens einmal vollständig durchgelesen und überarbeitet haben. Ebenso sollte man Artikel nicht aus dem Affekt schreiben, sondern sich Zeit nehmen, um mit vollem Menschenverstand einen Artikel zu verfassen.

### 2.3.6 Vernachlässigen Sie nicht die Aufmachung Ihres Artikels!

Es ist auf Groß- und Kleinschreibung zu achten, damit der Text leserlicher wird. Ebenso sollte der Artikel in sinnvolle Absätze untergliedert sein. Die Breite der Zeilen sollte unter 70 Zeichen liegen, da Zeilen über 70 Zeichen vom Menschen nur noch mit sehr großen Anstrengungen zu lesen sind. Generell sollte ein Artikel den Anforderungen der DIN 5008[3]-Norm entsprechen.

### 2.3.7 Achten Sie auf die "Subject:"-Zeile!

Die Subject-Zeile sollte eine kleine Inhaltsangabe des Artikels ( möglichst unter 40 Zeichen ) beinhalten, damit ein Leser schnell entscheiden kann, ob der Artikel für Ihn von Interesse ist, oder nicht, was bei der Vielfalt von Diskussionen eine richtige Erleichterung darstellt. Ebenfalls sollte man die Anrede nicht vergessen.

Ein kleines Beispiel:

Subject: Sind unsere Kinder zu groß ???

Hallo Herr Maier,
bezugnehmend auf Deinen Artikel möchte ich ...

### 2.3.8 Denken Sie an die Leserschaft!

Befindet man sich in einer internationalen oder regionalen Newsgroup? Oder in einer wissenschaftlichen Diskussionsgruppe oder nicht? Eine Gruppe in die man postet sollte man auch selbst lesen. Wenn möglich nur in eine Gruppe posten, nicht durch "Crossposting" mehrere Gruppen ansprechen.

### 2.3.9 Vorsicht mit Humor, Ironie und Sarkasmus!

In der elektronischen Kommuniaktion kann nichts durch Gestik und Mimik übermittelt werden. Deshalb ist darauf zu achten, dass ironisch gemeinten Bemerkungen gekennzeichnet werden, um keine Missverständnisse zu provozieren. Kennzeichnungsymbole sind die sogenannten Smilies ":-)" und ";-(".

### 2.3.10 Kürzen Sie zitierten Text auf das notwendige Minimum!

Zitate werden dem eigenen Kommentar vorangestellt und sollten nur den für den Kommentar relavanten Text umfassen. Beim richtigen Zitieren wird der Text durch das Zeichen >, bzw » eingerückt, um sofort ersichtlich zu machen, dass es sich hierbei um ein Zitat handelt. Zitieren der Signatur sollte unterlassen werden.

Ein Beispiel:

Herr Maier, schrieb am 11.11.00:
>Ich finde das unsere Kinder viel zu groß sind und
>zu viel essen ...
Finde ich auch, was auch die wissenschaftliche Studie XYZ zeigt. ...

[3]Schreib- und Gestalungsregeln für die Textverarbeitung

### 2.3.11 Benutzen Sie E-Mail!

Artikel in Newsgroups sollten nur solche sein, die von allgemeinen Interesse (Millionenpublikum) sein könnten, für alles andere, wie Hinweise auf Rechtschreibfehler, offensichtlicher Irrtümer oder Regelverstöße, ist eine E-Mail geeigneter.

### 2.3.12 Geben Sie eine Sammlung Ihrer Erkenntnisse an das Netz weiter!

Die Antworten auf Fragen, welche man per E-Mail empfangen hat, sollten sinnvoll zusammengefasst und veröffentlich werden, damit auch andere User, die ähnliche Fragen haben, eine Antwort finden können. Außerdem liest niemand gerne eine Newsgroup in der nur Fragen und keine Antworten stehen.

### 2.3.13 Achten Sie auf die gesetzlichen Regelungen!

Das Internet, also auch das Usenet ist kein rechtsfreier Raum, auch hier müssen Gesetze und Regelungen, wie das Urheberrechtgesetz eingehalten werden. Ebenso sollten E-mails von Dritten nicht veröffentlicht werden, ohne Zustimmung des Verfassers.

### 2.3.14 Benutzen Sie Ihren wirklichen Namen, kein Pseudonym!

Viele User agieren unter falschen Namen, was einige aufgrund Ihrer Anonymität zu beleidigenden und strafbaren Aussagen verleitet. Aufgrund diesen negativen Erfahrungen, die viele Leute im Netz mit den Trägern solcher Pseudonyme gemacht haben, sollten Autoren Ihre Artikel mit Ihrem wirklichen Namen versehen. Zeige Deine wahre Identität und stehe zu Deinen Aussagen.

### 2.3.15 Vorsicht mit Kommerziellem!

Das Usenet ist für den Informations- und Wissensaustausch geschaffen worden, und nicht für Kommerzielles.

### 2.3.16 Vorsicht mit Binaries und Multipart-Artikeln!

Binär-Dateien (Grafik, Musik, ausführbare Programme usw.) sind in den Diskussionsgruppen des Usenetzes unerwünscht. Da diese nur download-intensiv sind und somit den Usern Kosten verursacht. Sollte eine Binär-Datein unbedingt benötigt werden, sollte auf das WWW oder auf ftp-Downloadmöglichkeiten zurückgegriffen werden. Im deutschsprachigen Usenet wird auf aufwendige Textverarbeitungsprogramme verzichtet, deshalb sollen nur einfache Texte eingebracht werden.

### 2.3.17 "Du" oder "Sie".

Bei der Kommunikation zwischen fremden Menschen taucht immer die Frage auf, ob man duzt oder siezt. Generell ist es im Usenet sehr schwierig diese Frage zu beantworten, da man seinen Gegenüber meist nicht kennt. Deshalb wurde die Faustregel geschaffen, welche besagt: Wer selbst siezt, will gesiezt werden. Wer duzt, will selbst geduzt werden. Generell hat sich aber das "Du" etabliert.

### 2.4 Die Netiquette der Webseitenautoren

Die Hauptregel für Webseitenautoren besagt, dass Web-Seiten so gestalten werden sollen, dass möglichst jeder Benutzer, auch mit sehr leistungsschwachen Modems, angenehm surfen kann, d.h. nur das Nötigste an Grafik einbinden, und die Benutzer nicht dazu zwingen über Maps zu navigieren. Ebenso gehört zu jeder Webseite ein Vermerk auf die letztmalige Aktualisierung und auf den Verantwortlichen der Webpage.

## 3 Probleme, Diskussionen und Aktualisierungen der (.de*)-Netiquette

Zuerst stellt sich die Frage, wer bestimmt was in der Netiquette steht, bzw. stehen darf ? Wer hat das Recht die Netiquette zu ändern ?

Grundsätzlich baut die Netiquette auf dem Konsens des Netzes auf. Eine Änderung ist nur dadurch möglich, dass man den Maintainer davon überzeugt, dass sie sinnvoll und notwendig ist. Zu dieser Überzeugung kann auch eine Abstimmung in den thematisch geeigneten Gruppen sinnvoll sein.

Heutzutage nuzten immer mehr Menschen das Internet als Informationsquelle. Viele haben bisher nur die Dienste des WWW in Anspruch genommen, doch immer mehr Menschen wagen Sich jetzt auch ins Usenet oder wollen die Möglichkeiten des ftp-Downloads nutzen, was zu sehr vielen Regelverstößen führt. Auch immer neue Arten des Verstoßes werden bekannt, auch aufgrund ständiger technischer Erneuerungen. Daraus resulieren die Überlegungen vieler, ob man die Netiquette nicht aktualisieren sollte. Was gleichzusetzen ist mit einer Verschärfung des Regelwerks. Die Folge sind Metadiskussionen ( 600 Beiträge und mehr ) über mögliche Veränderungen der Netiquette, bei denen schon deutlich wird, dass eine Aktualisierung der Netiquette in nächster Zukunft sehr unwahrscheinlich ist.

**Die aktuellen Probleme und Diskussionsthemen sind:**

- Problem: Realnames

  Viele wollen im Schutze des Anonymität im Internet Ihren Schabernack treiben und sind nicht bereit Ihren Namen preiszugeben. Oftmals aber auch ohne strafrechtlichen Hintergrund, sondern heutzutage wird, z.B. bei Bewerbungen versucht alles über den Bewerber herauszufinden, dabei werden sogar seine Kommentare im Usenet durchforstet. Generell zu beachten ist aber, dass das Internet kein rechtsfreier Raum ist und man sogar zur Angabe des Realnamen gezwungen werden könnte.

- Problem: verfälschte E-Mail-Adresse

  Im Usenet gefundene E-mailadressen werden immer häufiger von Spammer missbraucht, deshalb verschlüsseln viele User Ihre Emailadressen, meist durch sinnlose Buchstabenkombinationen, wie Hans.xyzMaier@gmx.de. Doch einigen Usern ist es zu aufwändig, diese wieder zu entschlüsseln. Ein

weiterer Grund der gegen die Verschlüsselung spricht ist, dass dies auf lange Sicht kein wirksamer Spam-Schutz sein wird, denn es werden bald verschiedene Automatismen auf den Markt treten, welche die verschlüsselten E-Mailadressen wieder entschlüsseln können.

- Problem: HTML

  HTML bietet vielfältige Möglichkeiten wie verschiedene Schriften, Schriftgrößen, Farben und Textformatierungsmöglichkeiten, die manche gerne anwenden wollen. Aber dadurch würden sich auch die Ladezeiten erheblich erhöhen und die Übersichtlichkeit der Artikel verloren gehen.

- Problem: Binaries

  Die Frage, ob Binaries generell zugelassen werden sollen, drängt sich auf. Dafür spricht, dass man dadurch Sachverhalte schneller darstellen könnte (durch Grafiken), aber auch hier würde sich die Ladezeit erheblich erhöhen und die Gefahr der Vervirung des Usenets würde sich drastisch erhöhen.

- Problem: Duzen

  Einige wollen nicht geduzt werden, obwohl es sich so eigentlich eingebürgert hat, denn sie fühlen sich moralisch dazu gezwungen Fremde und gesellschaftlich "Höhergestellte" zu siezen. Aber das gesiezt werden interpretieren einige User als sehr beleidigend und störend.

Was aus den Beispielen deutlich wird, dass sehr viel über die Netiquette diskutiert wird, wobei sich die Fronten immer mehr verhärten.

# 4 Meine Version der (.de*)-Netiquette für deutschsprachige Newsgroups

Aus Banalität wurden einige Punkte gestrichen, da man von der Einhaltung dieser Regeln ausgehen kann. Ebenso habe ich eine Veränderung Ihrer Gewichtung (Reihenfolge) vorgenommen und versucht die Netiquette so allgemein zu halten, dass Sie leicht auf regionale Änderungen angepasst werden kann.

1. **Vergessen Sie niemals, dass auf der anderen Seite ein Mensch sitzt !**

   Wurde übernommen, als Hauptregel, die eigentlich die ganze Thematik der Netiquette wiederspiegelt.

2. **Vergesse auch nicht, dass Du selber nicht anonym bist, sondern ein identifizierbarer Mensch.**

   Neu, man kann sich nicht hinter dem Computer verstecken und denken mir passiert nichts. Man sollte sich diese Regel immer vor Augen halten, und bedenken, dass man für sein Handeln und Benehmen verantwortlich gemacht werden kann.

3. **Sei tolerant und hilfreich, vor allem zu neuen Netzbesuchern.**

   Neu, neue Netzbesucher sollte unter die Arme gegriffen werden, und nicht durch öffentliche Rügen abgeschreckt werden. Nicht gleich flamen[4], wenn ein Newbie[5] Netiquettenfehler begeht, sondern Ihn darauf aufmerksam machen und Verbesserungsvorschläge unterbreiten. Jeder neue User bereichert die Informationsvielfalt im Netz.

4. **Lesen Sie erst die allgemeinen Richtlinien oder Dokumentationen zum richtigen Verhalten und Umgang.**

   Neu, viele neue Netzbesuchern denken, dass sie sich recht gut mit Computern und Ihren Anwendungen auskennen, vergessen dabei aber das Computer nicht gleich Computer ist. Viele denken, dass Sie deshalb keine Regeln mehr lernen oder anschauen müssen, denn Sie kennen bereits alle Regeln. Das ist leider nicht so, deshalb sollte unbedingt vor Gebrauch einer Anwendung genernell die Bedienungsanleitung gelesen werden, so auch beim Usenet.

5. **Erst lesen, dann denken. Noch einmal lesen, noch einmal denken. Und dann erst posten!**

   Übernommen, da viele Artikel nur überflogen werden, und Kommentare im Affekt und deshalb teilweise zu impusiv gemacht werden, sollte sich jeder beim Durchlesen der Texte und beim Verfassen von Artikeln Zeit nehmen.

6. **Achten Sie auf die "Subject:"-Zeile**

   Übernommen, wichtige Regel zur schnellen Informationsfindung.

7. **Schreiben Sie nichts, was nicht mit dem eigentlichen Thema zu tun hat.**

   Neu, sehr häufig wird vom eigentlichen Diskussionsthema abgewichen. Sollte es sich nicht vermeiden lassen, dann ist eine neue Diskussionsrunde zu beginnen, in der dann das neue Thema diskutiert werden kann.

8. **Versuche, als Spamschutz, Deine E-mail-Adresse durch Hinweise in der Signatur zu verschlüsseln**

   Neu, damit wird den zahlreichen Diskussionen rund um die verfälschte E-mail-Adresse ein Ende gesetzt. Es wird zum Eigenschutz eines jeden User erlaubt, dass er seine E-mail-Adresse verschlüsseln darf. Indem eindeutige Hinweise in der maximal 4 Zeilen umfassenden Signatur gemacht werden, wie die E-mailadresse zu entschlüsseln ist.

9. **Vorsicht mit Humor, Ironie und Sarkasmus**

   Übernommen, da es sonst häufig zu Missdeutungen und Irrtümern kommen kann.

[4] zurechtweisende Kritik ausüben
[5] Anfänger im Usenet

10. **Zitiere nur das Nötigste, und fasse Dich kurz.**

    Zusammengefasst und übernommen, vor allem sollte richtig zitiert werden und nur soviel, wie ein neutraler Leser benötigt, um den Kommentar zu verstehen. Die Relation zwischen Kommentar und Zitat sollte höchstens 1:1 sein, ansonsten sollte man das Zitat kürzen.

11. **Halte Dich an gesetzliche Regelungen**

    Übernommen

12. **Verwende und Respektiere die Netiquette**

    Neu, die Anzahl an Netiquettengegnern nimmt immer mehr zu. Viele argumentieren, dass die Netiquette kein Gesetz ist, sondern nur eine Empfehlung, und man sie deshalb nicht beachten braucht. Die Netiquette exisitert und deshalb sollte man sich an die darin genannten Punkte und Regeln halten.

13. **Sei vorsichtig mit Abkürzungen**

    Neu, da die Menschen immer "schreibfauler" werden, werden immer mehr Abkürzungen verwendet. Generell sollten keine Abkürzungen verwendet werden, da sie einigen Usern nicht bekannt sind, wodurch das Lesen und Verstehen von Artikeln sehr erschwert wird.

Gestrichen wurde:

- **Teilen Sie etwas Neues mit!**

  Fällt unter die Regel "Erst lesen, dann denken. Nochmal lesen, noch einmal denken, und dann erst posten."

- **Ihre Artikel sprechen für Sie - Seien Sie stolz auf Sie!**

  und

- **Vernachlässigen Sie nicht die Aufmachung Ihres Artikels!**

  Beide Regel fallen weg, da ohne ein gewisses Niveau des Artikels ihn keiner lesen wird und infolgedessen keine sinnvollen Antworten gepostet werden. Deshalb versucht jeder Verfasser von Artikel automatisch einen möglichst fehlerfreien und niveauvollen Text zu schreiben.

- **Nehmen Sie sich Zeit, wenn Sie einen Artikel schreiben!**

  Wurde gestrichen, da auch diese Regel in der Regel: "Erst lesen, dann denken. Nochmal lesen, noch einmal denken, und dann erst posten." schon beinhaltet ist.

- **Denken Sie an die Leserschaft!**

  Sollte selbstverständlich sein und regelt sich automatisch, denn wenn in eine falsche Neuwsgroup gepostet wird, dann wird dort keiner eine sinnvolle Antwort geben können.

- **Benutzen Sie E-Mail!**

  Strittiger Punkt, aber E-mail sollt nur für Hinweise auf Verhaltensverstöße und Verständnisfragen benutzt werden, nicht aber für Antworten. Viele Mitlesende wollen auch die Meinung anderer verfolgen und nicht nur eine subjektiv zusammengestellte Ergebnissammlung durchlesen.

- **Geben Sie eine Sammlung Ihrer Erkenntnisse an das Netz weiter!**

  Entfällt, aufgrund der gestrichenen Regel: Benutzen Sie E-Mail!, denn wenn alle Antworten direkt in eine Newsgroups gepostet werden, braucht sich keiner mehr die Arbeit und Mühe zu machen, eine Ergebnisauswertung zu verfassen.

- **Benutzen Sie Ihren wirklichen Namen, kein Pseudonym!**

  "Normalbenutzer" des Usenetzes können nur sehr schwer nachprüfen, ob Hans Maier auch wirklich Hans Maier ist. Nur mit technisch aufwändigen Verfahren kann über die IP-Adresse der Name des Users ermittelt werden. Außerdem merkt man bald, dass man nur dann Antworten bekommt und ernst genommen wird, wenn man einen bürgerlichen Namen angegeben hat, über den man direkt angesprochen werden kann.

- **Vorsicht mit Kommerziellem!**

  Jeder weiß, dass über das Usenet keine Geschäft zu machen sind, deshalb versuchen nur noch wenige User Werbungen zu posten.

- **Vorsicht mit Binaries und Multipart-Artikeln!**

  Entfällt, denn mittlerweile werden solche Dateien über das WWW oder ftp angeboten, und wenn jemand nach Binär-Dateien sucht, weiß er wohin es sich wenden kann. Das Usenetz dient dem Austausch von Nachrichten und Informationen, worunter Binaries auch fallen können, also sollte das Einbetten von Binaries durchaus erlaubt werden.

  Multipart-Artikel sollten im Gegensatz nicht gepostet werden, da sonst die Übersichtlichkeit der Texte verloren geht, und damit die Informationsfindung erschwert wird.

- **"Du" oder "Sie"**

  Fällt weg, da es unter die neue Regel: Sei tolerant und hilfreich, vor allem zu neuen Netzbesuchern, fällt. Mit der Zeit kristallisiert sich heraus, dass der Eine gerne geduzt werden möchte und der Andere lieber mit Sie angesprochen werden will.

Eine weitere Überlegung ist ein *Tutorial* für die sogenannten Newbies einzuführen. Denn der mehrseitige Text über die Netiquette werden sich die wenigsten User durchlesen. Anstelle dieses Regelwerks sollte man lieber Schritt für Schritt eine Einführung in das Usenet geben und durch Beispiele auf mögliche Anfängerfehler aufmerksam machen. Auch die FAQ[6] in diesem Bereich sind so umfangreich, dass kein Mensch sie durchforsten wird, deshalb sollten auch hier die Fragen katalogisiert und aufs Wesentliche gekürzt werden.

[6] frequently asked questions

# 5 Ausblick

Im Gegensatz zu anderen Ländern, wie Amerika, halten sich in Deutschland relativ viele User an die Vorgaben der Netiquette, weshalb in deutschen Newsgroups auch noch sinnvoll diskutiert werden kann. Das sollte auch in Zukunft so bleiben! Dies geht aber nur wenn gewisse Regeln, wie die Netiquette eine ist, beachtet werden.

Zu erwähnen ist auch, dass die Netiquette nie topaktuell sein kann, denn dann müsste man sie sehr oft aktualisieren, wobei bald keiner mehr weiß welche Version die Aktuellste ist. Dies würde nur zu zusätzlichen Verunsicherungen der User, vor allem von Newbies, führen.

Zum Schluss noch ein Zitat von Andreas M. Kirchwitz, dem Verfasser der (.de*)-Netiquette, was die ganze Thematik rund um die Netiquette widerspiegelt.

"Was die Netiquette empfiehlt, ist so banal, daß selbst jeder Newbie von ganz allein darauf käme, würde er genauer darüber nachdenken. Die Netiquette ist die Basis. Sie sollte einen gewissen Grad an Zeitlosigkeit haben. Hielte sich jeder an die Netiquette, wäre schon viel gewonnen."[7]
(Maintainer Andreas M. Kirchwitz)

# 6 Literaturverzeichnis

Die vollständige Version der (.de*)-Netiquette findet man unter:
de.newusers.infos oder unter http://www.kirchwitz.de/~amk/dni/netiquette

Deutsche Version der RFC 1855:
http://www.ping.at/guides/netmayer

Englische Version der Netiquette erstellt von Arlene H. Rinaldi:
http://wise.fau.edu/netiquette/net/

Englische Version der RFC 1855 erstellt von Sally Hambridge, Intel Corporation aus Santa Clara:
http://www.dtcc.edu/cs/rfc1855.html

[7] Beitrag in der Newsgroups: de.admin.news.regeln zum Thema: Gültigkeit von Dokumenten vom 22.02.1999, Andreas M. Kirchwitz

# Standards

**Gerd Gröner**

gerd.groener@informatik.uni-ulm.de

*Zusammenfassung:* Standards und Normen sind in unserer heutigen Gesellschaft nicht mehr wegzudenken. Aber warum sind Standards so wichtig und wie entsteht ein Standard? Diese und weitere Fragen werden im folgenden behandelt. Der Schwerpunkt in dieser Ausarbeitung liegt bei Internetstandards, insbesondere die Entstehung und die Vorteile von Internetstandards.

## 1. Einleitung

Der große Boom des Internets entstand in den 90er Jahren, als auch Privatpersonen begannen das Internet zu nutzen. Danach gab es eine rasante Zunahme der Internetnutzer. Diese Zeit wird auch als Beginn des Informationszeitalters bezeichnet. Viele, hauptsächlich kleine und flexible Unternehmen erkannten rechtzeitig die Dynamik des Internets und versuchten schnell sich auf dem wachsenden Markt zu etablieren. Dadurch entstanden auch viele schnellexpandierenden Unternehmen. Das Internet wird von Personen aus der ganzen Welt benutzt, weltweite Kommunikation und Datenaustausch, dies kann nur geschehen, wenn es einheitliche Standards gibt, die Richtlinien für sämtliche Technologien festlegen.

Probleme entstehen bei der Umsetzung und Einhaltung von Standards. Es gibt wirtschaftliche und marktpolitische Anreize von Unternehmen, Standards nach ihren eigenen Vorstellungen und Wünschen durchzusetzen. Somit kommt es zu Konflikten zwischen verschiedenen Interessenvertretern eines Standards und es entsteht ein sogenannter Standardkrieg.

Es wird sich herausstellen, dass nur unabhängige und offene Standards eine breite Akzeptanz finden und somit eine weite Verbreitung garantieren. Solche Standards können nur durch unabhängige Gremien entstehen.

## 2. Grundlagen und Begriffe

Standards haben in der Industrie bereits eine lange Geschichte. Sie wurden in der Industrialisierung durch zunehmende Arbeitsteilung notwendig. Allgemeine Anforderungen an Standards sind: Einfachheit, leichte Umsetzbarkeit, Mächtigkeit

(große Akzeptanz) und Neutralität (keine Partei soll bevorzugt werden). Internetstandards beinhalten alle internetbezogenen Standards. Das sind Standards zur Darstellung und Präsentation von Webseiten, wie beispielsweise XML Standard, Standards die sich mit Datentransport befassen, zum Beispiel Email oder Protokolle. Kompatibilitätsstandards definieren Schnittstellen von verschiedenen Technologien und Produkte um eine Austauschbarkeit zu garantieren. Um die Etablierung von Standards zu verstehen, müssen auch die Begriffe Offenheit und Verbreitung betrachtet werden. Offenheit bedeutet die Offenlegung der verwendeten Technik und Methoden. Bei Programmen entspricht das dem Offenlegen des Quellcodes und der Implementierung. Offenlegung bedeutet automatisch freie Verfügbarkeit und die Möglichkeit selbst Modifikationen vorzunehmen. Das Gegenteil von Offenheit sind Eigentumsrechte oder Eigentumsansprüche. Bei immateriellen Gütern, beispielsweise Technologien werden Eigentumsrechte durch geistiges Eigentum begründet.

Es existiert auch ein Zusammenhang zwischen Offenheit und Verbreitung eines Standards. Das Ziel eines Standards ist immer eine hohe Verbreitung und eine größtmögliche Offenheit, dies führt dann zu einer großen Akzeptanz der Anwender.

# 3. Entstehung und Etablierung von Standards

In diesem Kapitel werden die verschiedene Möglichkeiten beschrieben, wie Standards entstehen können. Standards die durch Gesetzte und Gremien entstehen werden als de-jure-Standards bezeichnet. Standards die durch Marktführer entstehen werden als de-facto-Standards bezeichnet. Diese Entstehen aus internen (Firmeninternen) Standards, die sich auf dem Markt durchgesetzt haben.

## 3.1 Entstehung durch Gesetze

Diese Standards entstehen durch Gesetze oder durch Verordnungen öffentlicher Behörden, wie beispielsweise durch die Europäische Union. Diese Standards entstehen in der Regel vor Einführung der entsprechenden Produkte. Diese sind sehr ausgiebig und extrem detailliert und werden von den entsprechenden Herstellern nur soweit wie nötig umgesetzt. Diese Standards sind besonders für Verbraucher von Vorteil, da es keine Unsicherheit darüber gibt, welche Technologie sich durchsetzt. Der Vorteil für Anwender ist, dass diese Standards sehr konkret sind und den Herstellern wenig Gestaltungsfreiräume gewähren. Ein Beispiel in der EU sind Spezifikationen für ISDN und UMTS.

## 3.2 Entstehung durch Standardisierungsgremien

Es gibt viele nicht gewinnorientierte Organisationen, deren Aufgabe die Erstellung und Durchsetzung von Standards ist. Das Ziel dieser Gremien ist eine Maximierung

des Nutzen aus dem Standard aller Beteiligten, d.h. ein möglichst großer Gemeinnutzen und keine Gewinn- oder Profitinteressen einzelner Akteure.

### 3.2.1 Standardisierungsgremien

Die Mitglieder in Standardisierungsgremien sind hauptsächlich Vertreter aus Unternehmen und Wissenschaftler. Unter Wissenschaftler sind hier Fachleute gemeint, die keine eigenen, gewinnorientierten Interessen bei Standards verfolgen, sondern eher ideologisch und unabhängig ein möglichst hohen Gemeinnutzen mit einem Standard verfolgen. Bei den Mitgliedern aus Unternehmen ist es schwieriger, sie haben häufig einen Konflikt zwischen einem guten Standard, der einen hohen Nutzen für jeden bringt und andererseits den Gewinn- und Erfolgsinteressen des eigenen Unternehmens.

#### *3.2.1.1 W3C (World Wide Web Consortium)*

Das W3C [7,19] wurde 1994 an der MIT/LCS in Boston gegründet. Geleitet wird das W3C derzeit vom Direktor Tim Berners-Lee. Das W3C hat weltweit mehrere Niederlassungen, eine davon in Deutschland. Das W3C ist kein gewinnorientiertes Unternehmen.

Die Aufgabe des W3C besteht darin, Web Standards zu entwickeln. Das Hauptziel dabei ist es, das Web und auch die Vorteile des Webs für alle frei nutzbar zu machen[1]. Dies führt zu der schwierigen Aufgabe, stets einen Interessenausgleich zwischen den einzelnen Mitgliedern zu finden. Die Mitglieder sind überwiegend Software und Hardware-Hersteller, die natürliche daran interessiert sind, ihre eigenen Vorstellungen und Wünsche in den Standard mit einzubringen und gleichzeitig verhindern, dass sich konkurrierende Unternehmen bei einem Standard durchsetzen.
Die Kernaufgabe des W3C liegt darin, sogenannte Recommendations (=Empfehlungen) zu erstellen. Diesen Prozess, an dem alle Mitglieder beteiligt sind, sieht wie folgt aus:

- Working Draft
- Candidate Recommendation
- Proposed Recommendation
- Recommendations

Dieser Prozess wird innerhalb sehr kurzer Zeit durchlaufen, um einen schnellen Standardisierungsvorgang zu gewährleisten. Nur so kann sich ein Standard durchsetzen, wenn die Standardisierung annähernd parallel zur technischen Entwicklung der Unternehmen läuft.

Die Patentpolitik ist ein wichtiger Aufgabenbereich des W3C und wird von einer speziellen Patentarbeitsgruppe PAG (Patent Advisory Group) ausgeführt. Die PAG soll mögliche Probleme mit Lizenzen bei der Standardisierung lösen, falls Patente nicht mit den Lizenzanforderungen (insbesondere RF) vereinbar sind.

Es gilt stets der Grundsatz der „offenen Standards“. Dies bedeutet insbesondere für die Mitglieder, sämtliche Patente offen zu legen, aber auch eine kostenlose Implementierung von Empfehlungen des W3C. Derzeit gibt es zwei spezielle Lizenzierungsbedingungen für Webstandards, dies sind Royalty-Free (RF) und RAND (Reasonable and nondircriminatory). RAND erlaubt geringe und angemessene Patent- oder Lizenzgebühren für die Benutzung und Implementierung von Standards,

wobei RF ausdrücklich keine Gebühren für die Benutzung erlaubt, also wirklich freie Standards. Es gibt viele Befürworter für RF (z.B. HP, Netscape, Sun ...) und ebenso Gegner (z.B. MS). Eine Working Group (Arbeitsgruppe im W3C, bestehend aus Mitgliedern) kann bei jeweiligen Arbeit entscheiden ob RF oder RAND. Die Idee von W3C ist es, grundlegende, zentrale Standards RF zu lizenzieren, so dass eine Implementierung ohne Gebühren möglich ist, dagegen sind bei weitergehenden (bevorzugt high-level) Technologien oder Spezialisierungen durchaus Gebühren für die Implementierung vorstellbar, also RAND.
Bei RAND darf für die Implementierung eine Gebühr verlangt werden, allerdings muss diese angemessen sein und nicht diskriminierend. Es dürfen auch keine weiteren Bedingungen an den Gebrauch der Technologie gestellt werden.

#### *3.2.1.2 IETF (Internet Engineering Task Force)*

IEFT [3, 19] ist eine internationale Gemeinschaft, die Mitglieder sind Wissenschaftler und Unternehmen. Die IETF kümmert sich um Standards für Email und Protokolle. Die Hauptaufgabe besteht in der Weiterentwicklung des Internets und Internettechnologien. Die eigentlichen Aufgaben des IETF werden in spezialisierten Arbeitsgruppen erledigt. IETF ist keine gewöhnliche Standardisierungsorganisation sondern die Mitglieder sind nur freiwillige Beteiligte. Als Mitglied wird bezeichnet, wer auf einer Mailingliste einer Arbeitsgruppe steht und aktiv mitdiskutiert.

Zu Beginn war ein sehr großer Anteil der Mitglieder Akademiker und Wissenschaftler, die teilweise noch ideologische Interessen hatten. In der 80-er Jahren hab sich das Verhältnis zwischen Wissenschaftler und Unternehmen sehr zu Gunsten der Unternehmen entwickelt. Dies führte bei IETF zum gleichen Problem wie bei W3C, dass nun Mitglieder, insbesondere Mitglieder in Arbeitsgruppen aus konkurrierenden Unternehmen stammen und somit primär unternehmenseigene Interessen bei Standards verfolgen. Die IETF befasst sich auch mit der technischen Ausarbeitung von Standards. Der Standardisierungsprozess läuft mit sogenannten RFCs (request for comments) ab. Jeder kann einen RFC bei der IETF einreichen, allerdings stammen die meisten von einer Arbeitsgruppe. Eine Managementgruppe namens IESG (Internet Engineering Steering Group) entscheidet, ob über diesen RFC eine Diskussion eingeleitet wird. Diese Vorschläge eines RFCs werden dann veröffentlicht in den entsprechenden Arbeitsgruppen bearbeitet, überarbeitet und verbessert. Nach mehreren Verbesserungen und einer gewissen Wartezeit wird dann ein RFC als ein „öffentlicher“ Standard ausgewiesen. Anfangs gab es bei Standards kein geistiges Eigentum bzw. Eigentumsrechte. Seit 1996 werden bei Standards auch Eigentumsrechte berücksichtigt, dies geschah durch starken Druck von kommerziellen Mitgliedern. Es sind nun auch angemessene, nicht benachteiligten Lizenzgebühren erlaubt, RND (reasonable, non-discriminatory) genannt[19].

#### *3.2.1.3 ISO (International Institute for Standardization)*

Die ISO [5] wurde offiziell 1947 von 25 Staaten gegründet. Das Ziel war eine internationale Koordination und Vereinigung von Industriestandards. Eine Vorstufe

war die IEC (International Electrotechnical Commission) in der Elektroindustrie und die ISA (International Federation of the National Standardizing Associations) in anderen Industriebereichen. Standards werden bei der ISO durch ein Technisches Komitee (technical committees) TC entwickelt. Innerhalb eines Technischen Komitees gibt es dann noch Unterkomitees SC und Arbeitsgruppen (WG). Diese befassen sich mit der Ausarbeitung von Standards. ISO Standards sind freiwillig, werden aber normalerweise durch Marktanforderungen zur "Pflicht". Standards werden entsprechend den Marktanforderungen entwickelt.

In den Technischen Komitees (TC) und den Subkomitees (SC) entsteht ein Standard in folgenden Stufen [6]:

*Proposal Stage:* ein Vorschlag über einen neuen Standard wird diskutiert, der Vorschlag kommt meistens aus der Industrie.

*Preparatory Stage:* Eine Ausarbeitung über einen Standardvorschlag zu dem angegebenen Problem bzw. Sachverhalt (Proposal Stage) wird durch TC und SC erstellt.

*Committee Stage:* Verteilung des Entwurfs an beteiligte Komitees um Anmerkungen und Änderungsvorschläge zu äußern und schließlich zustimmen.

*Enquiry Stage:* nun werden die Entwürfe oder auch DIS (Draft International Standard) genannt, an alle ISO Mitglieder zur Kommentierung und Abstimmung verteilt. Der DIS gilt als akzeptiert und wird zum FDIS (Final Draft International Standard), wenn 2/3 aller am Standard beteiligten Mitglieder zustimmen und 3/4 aller ISO Mitglieder einverstanden sind.

*Approval Stage:* Der FDIS wird wieder an alle Mitglieder verteilt. Es gelten wieder die gleichen Abstimmungsbedingungen wie zuvor.

*Publication Stage:* Veröffentlichung des Standards.

### 3.2.2 Standardisierungsablauf am Beispiel vom W3C

Ein Standardisierungsprozess beim W3C besteht aus mehreren Phasen, in denen jeweils ein Dokument ausgearbeitet wird. Die wesentlichen Dokument beim Standardisierungsprozess sind:

- Submissions
- Working Draft
- Candidate Recommendation
- Proposed Recommendation
- Recommendation

Submissions sind die eigentlichen, konkreten Vorschläge von Mitgliedern für einen möglichen Standard. Dieser Vorschlag wird dann von Arbeitsgruppen mit weiteren Kommentaren und zusätzlichen Vorschlägen angereichert und dann als Zwischenergebnis (Working Draft) an Mitglieder veröffentlicht.

Bekommt ein Working Draft die Zustimmung aller Mitglieder der Arbeitsgruppe, so wird dieser Vorschlag als Proposed Recommendation bezeichnet. Dieser Vorschlag (Proposed Recommendation) geht dann an ein Komitee (Advisory Commitee), das diesen Vorschlag begutachtet. Das Advisory Commitee hat auch die Funktion eines Vermittlers zwischen Arbeitsgruppen und Mitgliedsorganisationen. Wird ein Vorschlag auch vom Advisory Commitee unterstützt, so wird daraus ein

Recommendation. Recommendations sind dann die eigentlichen Standards des W3C, sie stellen eine Einigung aller Mitglieder dar[1, 2, 7].

### 3.3 Entstehung durch Marktführer (Marktmacht)

Einige Standards entstehen nicht durch Gremien oder Gesetze, sondern durch Marktführer, bzw. durch Unternehmen mit besonderer Marktmacht, häufig Monopole. Standards, die so entstehen werden auch als de-facto-Standards oder informelle Standards bezeichnet. Ein de-facto-Standard entsteht dadurch, dass ein Unternehmen seine Vorstellungen und firmeninterne Standards bei seinen eigenen Produkten und Anwendungen benutzt. Hat ein Unternehmen eine entsprechend starke Marktposition dann werden diese Richtlinien oder ehemalige firmeninterne Standards von anderen Unternehmen und Anwendern einfach als Standard anerkannt. Dies ist natürlich nur bei entsprechender Marktmacht möglich. Allen anderen Unternehmen bleibt dann nichts anderes mehr übrig, als diese Standards anerkennen, um den eigenen Absatz nicht zu gefährden. Die Kompatibilität muss auch zu einem de-facto-Standard von allen anderen Unternehmen berücksichtigt werden, um Anwender nicht vom eigenen Produkt abzuschrecken. Sehr erfolgreich war mit de-facto-Standards Microsoft, beispielsweise mit Windows. Das Betriebssystem Windows ist seit einigen Jahren ein informelles Standardbetriebssystem.

Auch de-facto-Standards werden von anderen Unternehmen anerkannt, um den Vertrieb der eigenen Produkte zu sichern. So achten Softwarehersteller darauf, dass ihre Programme unter Windows Betriebssystemen laufen und eventuell mit Microsoft Office Programmen kompatibel sind. Wird dieser de-facto-Standard ignoriert, so hat der Softwarehersteller enorme Absatzprobleme mit den eigenen Produkten.

Interessant ist bei de-facto-Standards, dass diese nur entstehen, wenn ein Unternehmen ein deutlicher Marktführer ist. Bei zwei ungefähr gleich mächtigen Unternehmen oder Konkurrenten ist die Entstehung eines de-facto-Standards sehr unwahrscheinlich. Da sowohl Anwender als auch Hersteller auf anderer Produktionsstufe nur verunsichert werden, und möglichst versuchen, sich auf keinen der beiden potentiellen Standards festzulegen[24]. Die Unsicherheit der anderen Akteure verhindert so die Entstehung eines de-facto-Standards. Ein typisches Beispiel ist der Browserkrieg zwischen Netscape und Microsoft. Zwei gleich starke Konkurrenten führten einen Standardkrieg, Netscape dominierte zwar ursprünglich den Browsermarkt aber Microsoft war mächtiger durch die Verbreitung des Internet Explorers mit den Windows Betriebssystemen und durch Finanzierung der Weiterentwicklung des "kostenlosen" Internet Explorer mit anderen profitablen Microsoft Produkten. Keines der beiden Unternehmen gelang es, einen de-facto-Standard für Browser zu etablieren. Dagegen übernahm das W3C die unabhängige Standardisierung der Browsertechnologie. Es entstand somit ein de-jure-Standard. [25]

Oft werden allerdings auch Entstehungen von Standards durch Marktführer verhindert. Zum Beispiel wurde für das Grafikformat PNG bereits 1996 eine Empfehlung des W3C ausgesprochen, durchgesetzt hat sich aber PNG gegen GIF und JPG nicht. Der Hauptgrund war, da Microsoft mit dem Internet Explorer PNG-Formate nicht bzw. nicht richtig unterstützte[9].

Gelegentlich wird auch ein de-facto-Standard von einem Gremium nachträglich noch anerkannt. Das ist dann ein rein formaler Vorgang, kein eigentlicher Standardisierungsprozess.

### 3.4 Maßnahmen zur Durchsetzung eines Standards

In diesem Abschnitt wird der Weg beschrieben, um einen Standard zu etablieren. Hier wird das Verhalten der Anbieter betrachtet, also Unternehmen die diesen Standard durchsetzten und ihre Produkte und Anwendungen nach diesem Standard ausrichten. Es wird auch das Verhalten von Anwendern und konkurrierenden Unternehmen betrachtet. Hier werden unterschiedliche Ziele verfolgt. Eine möglichst weite Verbreitung bzw. eine große Bekanntheit eines Standards ist ein wichtiges Ziel. Ebenso wichtig ist die Offenheit oder Offenlegung von Standards. Hier entsteht häufig ein Zielkonflikt, da sich ein sehr offener Standard viel schneller verbreitet und anerkannt wird, aber für die Unternehmen geringere Erträge bedeuten, da es bei offenen Standards nicht möglich ist, Lizenzgebühren oder vergleichbare Abgaben zu verlangen.

Die Verbreitung von Standards ist nicht nur für potentielle Kunden bzw. Anwender wichtig, sonder es ist auch für bestehende Anwender nützlich, wenn eine Technologie möglichst weit verbreitet ist. Der Nutzen einer Technologie für den einzelnen steigt mit zunehmender Verbreitung, das bedeutet mit der Anzahl von Benutzern.

#### 3.4.1 Verhalten der Anbieter

Unternehmen die einen Standard durchsetzten müssen abwägen, ob sie eine große Verbreitung des Standards oder Gewinnerzielung aus Lizenzierung bzw. höheren Preisen (Monopol) von Eigentumsrechten bevorzugen. Der Optimalfall ist natürlich eine möglichst hohe Verbreitung und eine möglichst geringe Offenheit. Das ist allerdings schwierig zu erreichen, da bei wenig Offenheit die Ausbreitung geringer und deutlich langsamer ist. Dies entspricht dann einer monopolartigen Stellung, da geringe Offenheit ein hohes Maß an Eigentumsrechten bedeutet, also keine wesentliche Konkurrenz und gleichzeitig ein hoher Bekanntheitsgrad. Hier sind für Unternehmen die höchsten Erträge zu erzielen. Ein Beispiel für diese monopolartige Stellung ist wieder das Betriebssystem Windows.

Die Akzeptanz von Kunden und Anwendern steigt nicht nur mit Verbreitung sondern auch mit zunehmender Offenheit. Es existiert dann ein positiver Zusammenhang zwischen der Kombination Verbreitung Offenheit und der Akzeptanz der Kunden. Das bedeutet, um die Akzeptanz und somit indirekt auch die Verbreitung zu maximieren, ist es für ein Unternehmen vorteilhaft das Ziel der Offenheit und Offenlegung zu verfolgen.

Ein Beispiel ist Sun mit Java. Sun ist an einer starken Verbreitung von Java interessiert, unter anderem auch um von Microsoft Marktanteile zu erlangen. Sun spricht in diesem Zusammenhang vom "offenen" Java-Standard, um durch Offenlegung eine weite Verbreitung und Akzeptanz zu erzielen. Doch bei genauerer Betrachtung stellt sich heraus, dass die Standardisierung bei der ISO ausschließlich nach Vorstellungen von Sun erfolgte. Sun hat den Standard auch nie an die ISO abgegeben, sondern kontrolliert selbst die Entwicklungen in Java und besitzt die

Warenzeichenrechte von Java. Außerdem sind kommerzielle Java-Implementierungen ebenfalls kostenpflichtig. Dennoch verfolgt Sun das Ziel Offenheit um eine sehr weite Verbreitung zu erreichen.

### 3.4.2 Verhalten der Kunden bzw. Anwender

Anwender können die Durchsetzung eines Standards nur durch ihre Kaufentscheidung bzw. Benutzung beeinflussen. Den größten Einfluss haben die "normalen" Anwender, da sie einfach die Mehrheit bilden, also die Masse der Anwender. Die Akzeptanz eines Standards steigt mit zunehmender Verbreitung und mit zunehmender Offenheit. Für die Mehrheit der Kunden ist allerdings die Verbreitung wichtiger.
Innovationsfreudige und Freaks legen zwar eher Wert auf Offenlegung und freie Verfügbarkeit, diese Anwender stellen allerdings nur eine Minderheit dar.

Allerdings war in den letzten Jahren zu beobachten, dass der Marktanteil von Open Source Software deutlich zugenommen hat, also die Verbreitung von Open Source bzw. freier Software. Somit steigt auch für die meisten Anwender der Nutzen eines Produktes. Die zunehmende Verbreitung von Open Source Software wird durch die hohe Stabilität der Produkte begründet und natürlich auch aus finanziellen Aspekten. Unter anderem bei kleinen und mittelständischen Unternehmen, sowie bei öffentlichen Behörden. Grund hierfür ist hauptsächlich Kosteneinsparungen. Beispielsweise hat die Stadt Schwäbisch-Gmünd letztes Jahr auf Linux umgestellt. Der Grund waren die zu hohen Ausgaben für Lizenzgebühren für die bisher genutzten Microsoft Produkte.

### 3.4.3 Verhalten der Konkurrenz

Konkurrierende Unternehmen werden, je nach Einflussmöglichkeit, versuchen, einen Standard an dem sie nicht mitgewirkt haben zu vermeiden oder mindestens zu beeinflussen. Lässt sich ein Standard nicht mehr verhindern, so wird sich ein Konkurrent als letzte Möglichkeit schnell dem Standard anschließen, um weiteren Produktabsatz zu garantieren. Das bedeutet alle Produkte nach dem neuen, nicht gewollten Standard auszurichten. Das war beispielsweise bei IBM der Fall, nachdem sich das Betriebssystem OS/2 nicht gegen Microsoft Windows NT durchsetzten konnte.
Den größten Nachteil hat ein Unternehmen, wenn es versucht einen neuen, ungewollten Standard zu ignorieren. Dies führt zur fehlenden Kompatibilität zwischen den eigenen Produkten und Produkten anderer Unternehmen. Dann besteht die große Gefahr, dass das eigenen Produkt nicht mehr nachgefragt wird.
Kurz gesagt: ein Unternehmen wird immer versuchen, die Durchsetzung eines Standards von konkurrierenden Unternehmen zu verhindern, und eigen Vorstellungen durchzusetzen. Gelingt dies nicht, so wird das Unternehmen sich schnellstmöglich dem ungewollten Standard anpassen.

Ein Beispiel hierzu ist das Verhalten von Microsoft und Sun. Die plattformunabhängigen Java-Programme liefen anfangs unter Windows98 nicht, wenn diese für ein anderes Computersystem geschrieben waren. Laut Lizenzvertrag zwischen Sun und Microsoft müssen allerdings Java-Anwendungen unter allen

Betriebssystemen laufen. Die Änderungen an Windows98 wurden von Microsoft erst nach einer Klage durch Sun vorgenommen [20]. Allerdings hätte in diesem Fall Microsoft durch die Ignoranz des Java-Standards eher Sun geschadet als sich selbst.

## 4. Nutzen und Vorteile von Standards

Wozu braucht man überhaupt Standards? Wem nutzen Standards? Diese Fragen werden im folgenden näher betrachtet. Es gibt sowohl Vorteile für Hersteller als auch für Anwender.

Die Notwendigkeit von Standards kam erst durch zunehmende wirtschaftliche Verflechtungen und Arbeitsteilung in der Industrie. Durch die Aufteilung von Arbeitsprozessen ist es wichtig Normen und Richtlinien für Produkte und Produktabläufe zu beschreiben. Dies wird mit Schnittstellenkoordination bezeichnet [16]. Nach der Normierung von Produkten wurden auch Qualitätsstandards und Kompatibilitätsstandards festgelegt.

### 4.1 Kompatibilität

Es soll sichergestellt werden, dass Produkte von verschiedenen Herstellern problemlos gegenseitig ersetzt werden können. Dies ist sowohl für Anwender als auch für Hersteller ein wichtiger Nutzen von Standards. Anwender sind somit in der Lage, jederzeit Produkte von einem anderen Anbieter zu verwenden. Es ist ebenfalls ein großer Vorteil, wenn ein Produkt auch kompatibel mit der Vorgängerversion ist, das wird dann mit abwärtskompatibel bezeichnet. IBM machte beispielsweise 1987 einen Fehler durch mangelnde Abwärtskompatibilität der neuen Computer mit patentierten Bauelementen, um den Nachbauern der IBM-PCs Marktanteile wegzunehmen. Dies hatte allerdings zur Folge, dass die neuen IBM-PCs nicht mehr abwärts kompatibel zu den alten IBM-PCs und Nachahmer Produkten waren. Dadurch hat IBM deutlich Marktanteile verloren, hauptsächlich an Nachbauer der IBM-PCs.

Auch für Hersteller ist Kompatibilität wichtig. Hersteller auf nachgelagerter Produktionsstufe sind daran interessiert, die verwendeten Produkte bei verschiedenen Herstellern bzw. Lieferanten zu kaufen, ohne den eigenen Produktionsprozess anpassen zu müssen. Es ist ein hervorragendes Verkaufsargument, wenn ein Produkt kompatibel mit den Konkurrenzprodukten ist. Zum Beispiel hat Intel für bestimmte Baureihen von Prozessoren Lizenzen an AMD vergeben, um eine weite Verbreitung zu erreichen und Austauschbarkeit der beiden Prozessoren zu gewährleisten. Bei den leistungsstärkeren Nachfolgetypen wurde auf eine Abwärtskompatibilität zu den Vorgängern geachtet, um keine Marktanteile zu verlieren.

Ein konkretes Beispiel sind die Browser von Netscape und Microsoft. Anfangs konnte Netscape durch die starke Marktposition mit dem Navigator 2.0 den HTML-Standard beliebig festlegen. Dagegen hat Microsoft bei der zunehmenden Verbreitung des Internet Explorers eigene HTML Erweiterungen vorgenommen. Die verschiedenen Technologieerweiterungen wurden dann auf vielen Webseiten übernommen, dadurch bestimmte Features auf einigen Seiten entweder nur mit dem Internet Explorer oder nur mit dem Netscape Navigator zu sehen. Durch

Standardisierung soll hier erreicht werden, dass unabhängig vom Browser jede Seite gleich dargestellt wird.

Je genauer Standards sind, umso größer ist die Kompatibilität von Produkten, da es für die einzelnen Hersteller weniger Gestaltungsspielraum gibt.

### 4.2 Wettbewerb durch Standards – Wettbewerbssteigernde Wirkung

Durch Standard wird der Wettbewerb der Unternehmen deutlich verbessert[17, 18]. Hier spielen wieder die Kompatibilitätsstandards eine große Rollen. Dadurch werden die Markteintrittsbarrieren gesenkt. Typische Eintrittsbarrieren sind Patente, Urheberrechte und Gebrauchsmuster. Neue Anbieter können auf den Markt eintreten. Diese haben einen Kostenvorteil gegenüber den bisherigen Anbietern, die ihre hohen Entwicklungskosten in den Preis einkalkuliert haben. Dies fördert einen zusätzlichen Preis und Qualitätswettbewerb unter den Unternehmen. Insbesondere Anbieter von Freier Software können starken Einfluss auf kommerzielle Anbieter ausüben. Diese sind dann gezwungen durch bessere Leistung und Qualität den (höheren) Preis zu rechtfertigen. Ohne Standards könnten mächtige Unternehmen versuchen, Technologien geheim zuhalten und nur für gewünschte Produkte die Kompatibilität zu gewährleisten. Somit wäre ein Marktzutritt unmöglich.

Es gibt allerdings auch ein Problem damit. Durch die einfachen Marktzutritte und damit verbunden die geringere Rendite durch einen Preiskampf sinkt der Anreiz für Forschung und Entwicklung.

### 4.3 Qualität

Für Anwender soll die Qualität durch Standards verbessert werden. Dabei müssen die beiden Faktoren Unsicherheit und Neugier von Kunden betrachtet werden[16]. Die Unsicherheit veranlasst einen Kunden eher dazu, bekannte Produkte zu verwenden oder Produkte von namhaften, großen Herstellern. Billige oder kostenlose Produkte werden eher skeptisch bezüglich der Qualität und Leistung betrachtet. Ein Kauf und somit die Entscheidung für ein bestimmtes Produkt stellt immer ein gewisses Risiko dar, da ein potentieller Käufer vor dem Kauf einen Mangel an Information hat. Oft sind auch Vergleich von verschiedenen Produkten sehr schwierig. Andererseits ist ein Konsument auch neugierig auf andere, bisher unbekannte Produkte, die er gerne mal ausprobieren würde. Der Vorteil von Standards ist nun offensichtlich. Produkte sind besser vergleichbar, da Mindestanforderungen, gewisse Leistungsmerkmale und teilweise Qualität im Standard festgelegt sind. Ist auch eine Kompatibilität im Standard festgelegt, so besteht auch kein Risiko bei der Austauschbarkeit. Auch für "Laien" ist es ein Vorteil, wenn Produkte standardkonform sind. Sind die vom Kunden geforderten Eigenschaften im Standard definiert, so spart er sich mühsame Produktvergleiche, falls ihm die reine Erfüllung der Standardeigenschaften genügt.

### 4.4 Planungssicherheit

Besonders für Hersteller und Unternehmen ist dies ein großer Nutzen von Standards. Die Willkür eines Herstellers in einer anderen Produktionsstufe wird durch Standards verringert. Hält sich ein Hersteller an einen Standard, so können sich die anderen mit dem Hersteller agierenden Unternehmen auf diesen Standard einstellen. Der Hersteller wird bei Updates oder Folgeprodukten in der Regel weiterhin den Standard einhalten. Somit haben die anderen Unternehmen zumindest für die Grundeigenschaften der Produkte eine Vorstellung und somit eine bessere Planungssicherheit als ohne Standard.

## 5. Probleme der Standardisierung am Beispiel Royalty Free

Ein aktuelles Beispiel über Probleme, die bei Standardisierung auftreten ist Royalty Free (RF). Das konkrete Problem sind Patente auf offene Standards zu erheben, bzw. patentrechtlich geschützte Technologien in Web-Standards zu integrieren. Hier ist zwischen W3C und Microsoft ein Streit entstanden, ob dies erlaubt ist oder nicht. Der Begriff Royalty Free und die Bedeutung, also keine Lizenzgebühren auf Standards zu erheben, wurde bereits in Abschnitt 3.2.1.1 behandelt. Bei der Meinungsverschiedenheit sind auch andere W3C Mitglieder beteiligt, beispielsweise unterstützen Sun, HP, Apple und Netscape und viele Open-Source und free-software Anhänger die RF-only Politik von W3C, wogegen IBM die Position von Microsoft zumindest teilweise unterstützt.

### 5.1 Die Patent Politik von W3C

Die Grundidee ist, Internetstandards weitgehend mit lizenzfreien Methoden zu entwickeln, um grundsätzlich kostenlose Nutzung zu ermöglichen und somit eine weite Verbreitung zu gewährleisten. Das bedeutet grundsätzlich nur RF-Lizenzierung möglich, kurz RF-only genannt. Es gibt aber auch Ausnahmen, wenn es mit Patentinhabern keine Einigung gibt, so wird die geschützte Technik verwendet und die Anwender müssen dafür Lizenzgebühren bezahlen. Das soll aber nach der W3C Patent Politik Arbeitsgruppe nur eine Ausnahme sein, die meisten Standards sollen eine gebührenfreie Nutzung ermöglichen[22]. Falls für die Nutzung eines Standards Lizenzgebühren zu bezahlen sind, müssen Gebühren den RAND-Lizenzbedingungen des W3C entsprechen, dass bedeutet angemessene, verhältnismäßige und nicht diskriminierende Gebühren. Die Vereinbarung über RAND-Lizenzen werden dann zwischen W3C und Patentinhaber getroffen. Für die Lösung von Konflikten zwischen Lizenzinhaber und W3C Arbeitsgruppe ist innerhalb des W3C die PAG (Patent Advisory Group) zuständig [8]. PAG wird erst eingeschaltet, wenn in einer Arbeitsgruppe für eine RF-Standard festgestellt wird, dass eine patentierte Technologie verwendet wird.

Ein recht aktuelles Beispiel: Bei der Standardisierung von der Gestaltung von Web-Services und WSDL (Web Services Description Languages) werden zur Zeit zwei verschiedene Standards entworfen. Einerseits W3C mit WS-Choreography und

andererseits Microsoft, IBM und Bea Systems mit BPWL4WS. W3C versucht nun, den Standard in einer neutralen Syntax zu definieren, so dass beide mögliche Umsetzungen dem Standard entsprechen. Die Bedingung von W3C dafür ist allerdings: keine Lizenzgebühren für die Nutzung der Technik. Dadurch soll vermieden werden, dass sich gleichzeitig zwei verschiedene Standards entwickeln. IBM und Bea Systems sind mit dem Vorschlag von W3C einverstanden, Microsoft hat sich bisher noch nicht geäußert [10,11]. Das ist ebenfalls ein Standardkrieg, allerdings nicht zwischen zwei Unternehmen (wie zum Beispiel beim Browserkrieg) sondern zwischen Unternehmen und einem Standardisierungsgremium.

### 5.2 Argumente für Royalty Free

Unterstützung für Royalty Free bekommt W3C hauptsächlich von Open-Source Anhänger und Free-Software Anbieter aber auch viele kommerzielle Softwarehersteller. Die Royalty Free Unterstützer sind der Auffassung, dass die bisherige Patentpolitik des W3C Patentinhabern monopolistische Stellungen ermöglichte und sogar förderte, dagegen wurden open-source Anbieter, die nicht in der Lage sind, für ihre Produktentwicklungen Lizenzgebühren zu bezahlen, vom Markt verdrängt. Unterstützt wird RF unter anderem auch von HP, Sun und IBM.

Ein guter Standard ist dadurch gekennzeichnet, dass die verwendeten und beschriebenen Technologien kostenlos sind. Jeder soll in der Lage sein, einen Standard anzunehmen und Implementierungen danach ausrichten. Sind mit einem Standard Gebühren und Lizenzen verbunden, so wird dies kein guter Standard sein.

Die Legalität ist bei einem Standard wichtiger als die eigentliche Technik, die der Standard beschreibt. Dies ist das Hauptargument für freie Standards. Jeder der seine Implementierung an einem Standard ausrichtet, soll dafür Rechtssicherheit haben, ohne Lizenzgebühren bezahlen zu müssen. Es soll auch niemand, der sich an einem Standard orientiert, in Versuchung geführt werden, Lizenzen und Patente zu ignorieren und somit gesetzeswidrig zu handeln.

Microsoft wird vorgeworfen, wegen teilweiser Monopolstellung von ungelösten und nicht definierten Webstandards zu profitieren[14]. Dann ist Microsoft in der Lage, selbst die Richtlinien festzulegen, nach denen sich alle anderen richten müssen. Wenn sich Microsoft aktiv an einer Standardisierung beteiligt, dann nur um selbst möglichst viel Geld zu verdienen, auch auf Kosten des W3C. Der eigene Profit ist Microsoft wichtiger als gute Webstandards. Die Mitgliedschaft im W3C ist nur dadurch zu begründen, dass Microsoft seine eigenen Interessen bei der Standardisierung durchsetzen kann. Aus Sicht vieler W3C Mitglieder hat Microsoft was gegen freie Software.

### 5.3 Argumentation gegen Royalty Free

Die folgenden Argumente [14] gegen RF stammen von Microsoft, einem bekennenden Gegner des W3C Vorschlags. Die W3C RF-only Patentpolitik wird niemals zu nützlichen Standards führen. Es wird nur ein Kampf gegen Softwarepatente und kommerzielle Software gestartet. Das hat nichts mit guten

Standards zu tun und wird keine große Akzeptanz bewirken. Die Aufgabe von W3C ist es, Webstandards zu erhalten, ausbauen und erstellen, und nicht eine Interessenvertretung der free-software Anhänger zu sein. Die Effektivität von Softwareentwicklung hängt auch von Lizenzen und Patentgebühren ab. Es müssen Leistungsanreize existieren. So gesehen haben rein freie Standards eine wettbewerbsbeeinträchtigende Wirkung.

Microsoft hat schon an vielen guten Standards des W3C mitgewirkt. Beispielsweise mit XML-Data, Soap und WSDL.

Hat ein Unternehmen (ob W3C Mitglied oder nicht) Patente, die in einem Standard beschrieben werden, so kann das Unternehmen nach RF-only Patentpolitik keine Lizenzgebühr verlangen. Die einzige Möglichkeit um den Standard zu ermöglichen ist eine Freigabe des Patents, oder der Standard kann in der Form nicht realisiert werden. Dies kann zu langen Blockaden führen. Somit wird die Standardisierung nicht einfacher sondern bedeutend schwieriger.

Sind für einzelne Verfahren oder Techniken die zu einem Standard gehören Lizenzgebühren erlaubt, so kann dies laut Microsoft die Qualität des Standards erheblich verbessern. In diesem Fall kann für den Standard die beste Lösung bzw. die beste Methode verwendet werden, auch wenn das gelegentlich Patentgebühren zu Folge hat. Die Entscheidung für eine Technik kann dann nur nach qualitativen Gesichtspunkten erfolgen. Dürfen dagegen nur freie Techniken verwendet werden, so gibt es weniger Alternativen und viele gute Technologien können überhaupt nicht verwendet werden.

Laut Microsoft sind Standards nur erfolgreich, wenn auch Anforderungen und Wünsche von Kunden berücksichtigt werden. Viele Microsoft Kunden vor allem in der Elektro- und Telekomunikationsindustrie sind selbst W3C Mitglieder und besitzen viele Patente. Diese Kunden sind nach Aussage von Microsoft nicht bereit unter RF-only Patentpolitik an Standardisierung mitzuwirken. Allerdings hat sich noch keiner dieser Kunden direkt beim W3C zu Wort gemeldet.

Microsoft kritisiert die unterschiedliche Bewertung von Patenten bei Produkten und Techniken. Ein Microsoft Mitarbeiter erwähnt folgendes Beispiel: Jeder kann verstehen, wenn auf DVDs oder auf einen Web-Browser eine geringe Patenttantieme erhoben wird, aber wie soll beispielsweise auf HTML eine Patentgebühr erhoben werden.

## 6. Zusammenfassung und Bewertung.

Die Standardisierung erfolgt zunehmend durch Gremien. Auch die Bedeutung von Standardisierungsgremien hat deutlich zugenommen. Es ist auch erkennbar, dass sich Unternehmen zunehmend an die Empfehlungen der Standardisierungsgremien halten, und nicht nur unternehmensinterne Interessen durchsetzen wollen. Es wurde zwar in einer Studie vom Februar 2002 festgestellt, dass nur ca. 4 % der Web-Seiten dem W3C Standard entsprechen, und bei den meisten W3C Mitgliedern sind die Web-Seiten nicht standardkonform. Auch einige Browser und Autorenwerkzeuge sind

nicht standardkonform, wie beispielsweise von Herstellern Microsoft, AOL, Apple und Adobe[12]. Dennoch ist ein Trend zur deutlichen Verbesserung erkennbar.
Ein großer Erfolg des W3C war die Übernahme der technischen Führerschaft beim Browser-Krieg zwischen Netscape und Microsoft[25].

Ein weiterer und sehr wichtiger Trend ist die Zunahme von offenen und unabhängigen Standards. Nur diese Standards werden langfristig bestehen, da sie eher langlebig konzipiert sind und der Lebenszyklus nicht von einem Hersteller beeinflusst werden kann. Auch die Erweiterbarkeit, die erst durch Offenlegung des Quell-Codes möglich ist, spricht für offene Standards. Dieser Trend wird auch von Open Source Software gefördert. Open Source Software ist nur mit offenen Standards möglich, mit Technologien die ohne Lizenzgebühren verwendet werden können. Die Bedeutung von Open Source Software ist vielen Menschen nicht klar. Open Source Software ist nicht nur was für ein paar "Spinner" oder Freaks. Die meisten Leute arbeiten täglich, zum Teil unbewusst, mit Open Source Software. Sehr bekannte und beliebte Open Source Software sind beispielsweise der Apache Webserver (Marktanteil im Jahr 2000: 63 %), das Serverprogramm BIND (selbst auf Microsoft NT/2000 Servern), Sendmail (Mailtransfer), Perl (nicht-statische Webseiten, DB-Anbindung) und das Betriebssystem Linux.[21]

## 7. Literatur


[1] World Wide Web Consortium (W3C), www.w3.org/2001/08/16-PP-FAQ

[2] W3C Patent Policy Working Group, www.w3.org/2001/ppwg

[3] Internet Engineering Task Force, www.ietf.org/overview.html

[4] Internet Engineering Task Force, www.ietf.org/rfc/rfc2727

[5] International Institute for Standardization (ISO), www.iso.org

[6] ISO, www.iso.org/eu/about/iso/indroduction/index.html

[7] World Wide Web Consortium, www.w3.org

[8] W3C Patent Politik, www.w3.org/TR/2002/WD-patent-policy-20021114/

[9] "Patent für GIF-Grafikformat läuft aus", Artikel in Computerwoche, 03.07.03

[10] "Brücke zwischen WS-Choreagraphy und BPEL4WS", Artikel in Computerwoche, 31.01.2003

[11] "W3C promotes Royalty Free web services Standards", Artikel in InfoWorld News, www.infoworld.com, 14.11.2002

[12] "W3C Mitglieder halten sich nicht an die eigenen Standards", Computerwoche, www.computerwoche.de, 03.07.2003

[13] "Standards, Lügen und Microsoft", Artikel in c't 14/97, www.heise,de/ct, 1997

[14] Veröffentlichte Mails zu Royalty Free, gesendet an www-patent-policy-comment@w3.org, www.w3.org

[15] Offene Systeme und Standards, Stempfle, Diplomarbeit, 1996 www.fh-augsburg.de/informatik/diplomarbeiten/langfassungen/stempfle-stork-1996/opensystems/stds_1.htm

[16] "Herausforderungen systemtechnologischer Vertrauensgüter an das Marketing-Management", Backhaus & Späth , 1994

[17] "Bedeutung von Standards für Multimedia-Wertschöpfung", Piscitello, Seminararbeit

[18] "Wettbewerb, Standardisierung und Regulierung beim digitalen Fernsehen", Nicklas, 2000

[19] The Web Standard Project, www.webstandards.org

[20] "Internet Standards", Washtenaw Community College, http://courses.wccnet.org/computer/mod/ng07c.htm

[21] Freie Bürger fordern freie PCs, SPLITTER, IT-Nachrichten für die Berliner Verwaltung Nr. 1-2001, 2001, www.lit.verwalt-berlin.de

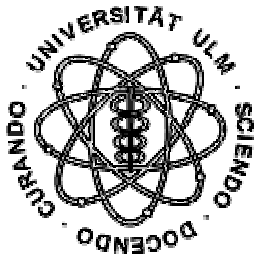
[22] "W3C regelt Umgang mit Patenten", Artikel in Computerwoche, 20.03.2003

[23] "W3C schlägt neuen Umgang mit Patenten vor", Artikel in Computerwoche, 29.01.2002

[24] "Unterschiedliche Browser-Technik lässt Entwickler verzweifeln", Artikel in Computerwoche, 12.04.1999

[25] "Standards – was kommt, was bleibt?", Klaus Birkenbihl, W3C, www.w3c.de

**Universität Ulm**
**Fakultät für Mathematik und Wirtschaftswissenschaften**




## Seminararbeit zum Seminar „Internetdienste“

# Thema: „Rechtliche Aspekte im Internet – Grundlagen des Rechts“

**Von Martin Entenmann**
**Wirtschaftswissenschaften, 8. Semester, Schwerpunkt Informatik**

# Inhaltsverzeichnis

1. Einführung....3
2. Das Urheberrecht....4
2.1 Der Urheber und seine Rechte....4
2.2 Ausnahmen des Urheberrechts....6
2.3 Haftung für Verletzungen des Urheberrechts im Internet....7
2.3.1 Haftung des Anbieters....7
2.3.2 Haftung des Nutzers....8
2.3.3 Haftung des Providers....8
3. Das Kennzeichenrecht....9
3.1 Allgemeines....9
3.2 Kennzeichenrecht und Internet-Domains....11
4. Datenschutz im Internet....14
4.1 Allgemeines....14
4.2 Datenschutzrechtliche Pflichten der Unternehmen....16
Literatur- und Quellenverzeichnis....18

# 1. Teil: Grundlagen des Rechts

## 1. Einführung:

Das Internet gilt als das Zukunftsmedium schlechthin. In der öffentlichen Diskussion steht es als Symbol für technischen Fortschritt, kulturelle Globalisierung und wirtschaftliches Wachstum. Die Marktpotenziale und Möglichkeiten des Netzes wachsen weiter - unabhängig von den Ups und Downs des wirtschaftlichen Umfeldes. Auch die Zahl der Nutzer steigt beständig. Mit dem Boom, den die Entwicklung des Internet in den letzten Jahren erlebt hat konnten jedoch weder die Rechtswissenschaft an sich, noch die Gerichte auch nur annähernd Schritt halten. Ein einheitliches Internetrecht als solches gibt es (noch) nicht. Statt dessen gibt es ein Sammelsurium von Gesetzen, rechtlichen Regelungen und Vorschriften, die auf das Internet übertragen und angewandt werden müssen. So gilt es z. B. die Rechtsvorschriften in Zusammenhang mit dem allgemeinen Vertragsrecht, also das Bürgerliche Gesetzbuch (BGB) und das Handelsgesetzbuch (HGB) zu beachten. Aber auch schwierigere Rechtsgebiete wie z. B. das Urheberrecht, das Kennzeichenrecht oder das Datenschutzrecht kommen hinzu. In engem Zusammenhang mit diesen Rechtsgebieten stehen dann beispielsweise auch noch das Teledienstedatenschutzgesetz, der Mediendienstestaatsvertrag oder die Telekommunikations-Datenschutzverordnung.
Ziel dieser Seminararbeit ist es, einige der weiter oben erwähnten Gesetze, rechtlichen Regelungen und Vorschriften etwas näher zu beleuchten, und sie anhand konkreter Urteilsfindungen zu verdeutlichen.

(Vgl. hierzu Prof. Dr. Wolfram Gass: Skript zur VL „Medienrecht“, Uni Ulm, 2003, S. 19f)

## 2. Das Urheberrecht:

Einführend muss festgestellt werden, dass der Umfang des Urheberrechts von Land zu Land unterschiedlich ausfällt. Daher ist der Standort des Werkes von entscheidender Bedeutung für den ihm gewährten Schutz. Oftmals wird das Urheberrecht mit dem Begriff „Copyright“ bzw. dem Symbol dafür („©“) gleichgesetzt oder zumindest mit ihm in Zusammenhang gebracht. Unter dem Copyright wird das us-amerikanische Urheberrecht verstanden. Dieses Copyright und das deutsche Urheberrecht sind sich in mancher Hinsicht ähnlich, jedoch keinesfalls identisch. Beim Copyright wird vor allem der ökonomische Aspekt in den Vordergrund gestellt, wobei die wirtschaftliche Investition sowie die produzierte Ware (copies) geschützt werden soll, nicht jedoch der Schöpfer des Werkes belohnt werden soll. Im Gegensatz dazu schützt das deutsche Urheberrecht die ideellen und materiellen Interessen des Urhebers an seinem Geisteswerk. Die folgende Arbeit bezieht sich jedoch rein auf das in Deutschland geltende Recht.

Gemäß § 1 des Urhebergesetzes (UrhG) werden Werke der Literatur, der Wissenschaft und der Kunst urheberrechtlich geschützt. Hierbei können sich auch durchaus Überschneidungen ergeben. Die Abstufung in die o.g. Begriffe ist wichtig im Hinblick auf die Abgrenzung zu anderen Rechtsbereichen, wie z. B. dem Patentrecht oder dem Geschmacksmusterrecht. Das deutsche Urheberrecht erfordert keine Anmeldung (wie z. B. das Patentrecht), sondern tritt automatisch unmittelbar nach der Schaffung des Werkes ein. Gemäß § 64 UrhG erlischt das Urheberrecht erst 70 Jahre nach dem Tod des Urhebers.

### 2.1 Der Urheber und seine Rechte:

Der Urheber wird laut § 7 UrhG definiert als der Schöpfer des Werkes, wobei nur eine natürliche Person Urheber sein kann. Unwichtig ist dabei, ob der Urheber im Sinne des BGB als geschäftsfähig gilt. Der Fall der Miturheberschaft tritt ein, wenn mehrere natürliche Personen gemeinsam ein Werk geschaffen haben, ohne dass sich ihre Anteile gesondert verwerten lassen. Die Erträge aus der Nutzung des

Werkes gebühren den Miturhebern nach dem Umfang ihrer Mitwirkung an der Schöpfung des Werkes, wenn nichts anderes zwischen den Miturhebern vereinbart ist (§ 8 UrhG). Haben mehrere Urheber ihre jeweils gesondert verwertbaren Werke mit dem Ziel einer gemeinsamen Verwertung miteinander verbunden, so spricht man gemäß § 9 UrhG von der Urheberschaft verbundener Werke. Dies wäre z. B. dann der Fall, wenn ein Urheber den Text, ein anderer die Grafik oder die Photografie zu einem Gesamtwerk beisteuern. In manchen Fällen ist nicht auf Anhieb ersichtlich, wer nun der Urheber eines Werkes ist. Hier gilt bis zum Beweis des Gegenteils derjenige als Urheber, der auf den Vervielfältigungsstücken eines erschienenen Werkes oder auf dem Original eines Werkes in der üblichen Weise als Urheber bezeichnet ist. Diese Vermutung der Urheberschaft ist in § 10 des Urhebergesetzes geregelt. Dies gilt auch dann, wenn der Urheber, Decknamen oder Künstlerzeichen für die Bezeichnung seines Werkes verwendet.

Das Urheberrecht schützt den Urheber in seinen geistigen und persönlichen Beziehungen zum Werk und in der Nutzung des Werkes (§ 11 UrhG). Daraus ergeben sich zum einen die sog. **Urheberpersönlichkeitsrechte** und zum anderen die sog. **Verwertungsrechte** des Urhebers.

Zu den **Urheberpersönlichkeitsrechten** zählen das z. B. das Veröffentlichungsrecht (§ 12 UrhG), nach dem der Urheber das Recht hat zu bestimmen, ob und wie sein Werk zu veröffentlichen ist. Des weiteren zählt die Anerkennung der Urheberschaft (§ 13 UrhG) dazu: Der Urheber hat das Recht auf Anerkennung seiner Urheberschaft am Werk. Er kann bestimmen, ob das Werk mit einer Urheberbezeichnung zu versehen und welche Bezeichnung zu verwenden ist. Er hat hier z. B. die Wahl zwischen seinem bürgerlichen Namen, einem Pseudonym, einem Künstlernamen usw.

Aus den **Verwertungsrechten** des Urhebers ergibt sich, dass der Urheber das ausschließliche Recht hat, sein Werk in körperlicher Form zu verwerten. Hierzu sind vor allem das Vervielfältigungsrecht (§ 16 UrhG), das Verbreitungsrecht (§ 17 UrhG) und das Ausstellungsrecht (§ 18 UrhG) zu nennen. Nach dem Vervielfältigungsrecht ist der Urheber berechtigt, Vervielfältigungsstücke des Werkes herzustellen, unabhängig vom Vervielfältigungsverfahren und unabhängig von der Vervielfältigungszahl. Das Verbreitungsrecht räumt dem Urheber das Recht ein, sein Originalwerk oder Vervielfältigungsstücke des Werkes der Öffentlichkeit anzubieten (Aufforderung zum Eigentum- oder Besitzerwerb des Stückes) oder in Verkehr zu

bringen. Das Ausstellungsrecht ist das Recht des Urhebers, sein Originalwerk oder Vervielfältigungsstücke eines unveröffentlichten Werkes öffentlich zur Schau zu stellen.

Des weiteren ist der Urheber natürlich auch berechtigt, sein Werk in „unkörperlicher Form“ öffentlich wiederzugeben. Dieses Recht umfasst insbesondere das Vortrags-, Aufführungs- und Vorführungsrecht (§ 19 UrhG), das Senderecht (§ 20 UrhG) sowie das Recht der Wiedergabe durch Bild- oder Tonträger (§ 21 UrhG). Das Vortrags-, Aufführungs- und Vorführungsrecht erlaubt dem Urheber sein Werk durch persönliche Darbietung zu Veröffentlichen. Das Senderecht gibt dem Urheber das Recht das Werk durch Funk (z.B. Ton- und Fernsehrundfunk, Satellitenrundfunk, Kabelfunk) oder ähnliche technische Mittel, der Öffentlichkeit zugänglich zu machen. Das Recht der Wiedergabe durch Bild- oder Tonträger erlaubt dem Urheber Vorträge oder Aufführungen des Werkes mittels Bild- oder Tonträger öffentlich wahrnehmbar zu machen.

Diese o. g. Verwertungsrechte werden als sog. ausschließliche (absolute) Rechte bezeichnet. Dadurch soll es dem Urheber ermöglicht werden, gänzlich über die Nutzung seines Werkes zu bestimmen sowie durch die entsprechende Nutzung zu profitieren.

## 2.2 Ausnahmen des Urheberrechts:

Ausgenommen vom Urheberrecht sind laut § 5 UrhG **amtliche Werke** wie z. B. Gesetze, Verordnungen, amtliche Erlasse und Bekanntmachungen. Es ist jedoch zu beachten, dass für amtliche Werke ein Änderungsverbot sowie die Verpflichtung zur Quellenangabe gilt.

**Öffentliche Reden**, die vor staatlichen, kommunalen oder kirchlichen Organen gehalten worden sind dürfen vervielfältigt und verbreitet werden (§ 48 UrhG).

Zulässig ist weiterhin die Vervielfältigung und Verbreitung **einzelner Rundfunkkommentare** und **einzelner Artikel** aus Zeitungen, sofern diese politische, wirtschaftliche oder religiöse Tagesfragen betreffen und nicht mit einem Vorbehalt der Rechte versehen sind (§ 49 UrhG).

Erlaubt ist auch gemäß § 51 UrhG das **Zitieren**: Darunter wird die Vervielfältigung, Verbreitung und öffentliche Wiedergabe eines urheberechtlich geschützten Werkes

auch ohne Zustimmung des Urhebers verstanden, wenn dieses in einem eigenständigen wissenschaftlichen Werk zur Erklärung des Inhalts aufgenommen wird. Wichtig ist hierbei, dass der Zitatsumfang wesentlich begrenzt wird.
Zulässig ist auch die **öffentliche Wiedergabe** eines erschienenen Werkes, wenn die Wiedergabe keinem Erwerbszweck des Veranstalters dient, und das Publikum unentgeltlich beiwohnen darf.
§ 53 UrhG sieht vor, dass **Vervielfältigungsstücke eines Werkes zum privaten Gebrauch** herzustellen zulässig ist. So dürfen beispielsweise Werke aus dem Internet für wissenschaftliche Zwecke kopiert und verwertet werden. Dies gilt nicht nur für Wissenschaftler, sondern auch für Forschungsinstitute, Studenten im Rahmen ihrer Ausbildung und auch für Privatleute mit wissenschaftlichem Informationsbedürfnis. Wichtig hierbei ist jedoch, dass der Urheber gem. § 54 UrhG einen Anspruch auf Zahlung einer angemessenen Vergütung hat.

## 2.3 Haftung für Verletzungen des Urheberrechts im Internet:

### 2.3.1 Haftung des Anbieters:

Da Verträge zwischen dem Provider uns seinem Nutzer immer eine Klausel enthalten, die besagt, dass keine Programme, Dateien etc., welche die Rechte Dritter verletzen in irgendeiner Weise angeboten werden dürfen, kommt durch das urheberrechtswidrige Anbieten von Dateien im Internet eine Verletzung dieses Vertrages zustande.
Daraus ergibt sich für den Provider gegenüber dem Anbieter urheberrechtswidriger Inhalte sowohl ein Kündigungsanspruch, als auch unter Umständen ein Schadensersatzanspruch wegen Pflichtverletzung. Gem. § 97 UrhG kann der Anbieter vom Urheber auf Beseitigung der Beeinträchtigung, bei Wiederholungsgefahr auf Unterlassung und, wenn dem Anbieter Vorsatz oder Fahrlässigkeit zur Last fällt, auch auf Schadensersatz in Anspruch genommen werden. Nutzt der Anbieter ein urheberrechtlich geschütztes Werk, so muss er sich selbständig über die Einwilligung des Urhebers Klarheit verschaffen.

## 2.3.2 Haftung des Nutzers:

Da der Nutzer in keiner vertraglichen Beziehung zum Urherber steht kommt eine vertragliche Haftung hier in der Regel nicht vor. Hat der Nutzer lediglich eine Vervielfältigung zum privaten und sonstigen eigenen Gebrauch gemacht, so kann er dafür auch nicht belangt werden. Hat er die Kopie jedoch weiterveräußert, oder ist sein Interesse an der Kopie von sonstiger kommerzieller Natur, so besteht in jedem Falle ein Schadensersatzanspruch des Rechteinhabers gegen ihn (§ 97 UrhG). Des weiteren hat der Urheber laut § 98 UrhG einen Anspruch auf Vernichtung oder Überlassung der Vervielfältigungsstücke.

## 2.3.3 Haftung des Providers:

Für den Provider ist es sehr schwierig in den großen gespeicherten Datenmengen urheberrechtswidrige Inhalte aufzuspüren. Es kann darüber gestritten werden, wie viel Aufwand zum Auffinden solcher Inhalte sinnvoll und zumutbar ist. Fest steht jedoch, dass der Provider gemäß der „notice and take down procedure“ bei einem konkreten Hinweis auf urheberrechtswidrige Inhalte selbige umgehend zu löschen hat. Bei Nichtbeachtung dieses Prinzips besteht ein Schadensersatzanspruch des Urhebers gegenüber dem Provider gem. § 97 UrhG. Dies gilt ebenfalls, wenn der Provider dem Anbieter Speicherplatz zur Verfügung stellte, obwohl er hätte erkennen müssen, dass der Anbieter damit gegen Urheberrecht verstößt.

Quellenangaben zu diesem Kapitel:

Deutsches Urhebergesetz

Gesellschaft zur Verfolgung von Urheberrechtsverletzungen e.V.: www.gvu.de/deutsch/fr_recht.html

Kanzlei Panke und Partner: www.panke-eurojuris.de/rechtsgebiete/urheber.html

Prof. Dr. Wolfram Gass: Skript zur VL „Medienrecht“, Uni Ulm, 2003, Teil 3

# 3. Das Kennzeichenrecht:

## 3.1 Allgemeines:

Ein Kennzeichen soll der Identifizierung und der Unterscheidung dienen. Das Kennzeichen unterscheidet also Personen bzw. Gegenstände von anderen Personen bzw. Gegenständen und ordnet einen Gegenstand seiner Herkunftsquelle (z.B. Betrieb oder Region) zu. Dem Kennzeichenrecht liegt ein dreifacher Interessenschutz zu Grunde. Der erste Interessenschutz ist der Schutz der wirtschaftlichen Leistung gegen Ausbeutung und Behinderung (also das Interesse der Gewerbetreibenden). Der zweite Interessenschutz ist der Schutz vor Irreführung (Interesse der Verbraucher). Der dritte Interessenschutz ist der Schutz der Marktordnung und der Wettbewerbsfreiheit (Interesse der Allgemeinheit). So kann sich z. B. der Abnehmer (Kunde) nur dann frei entscheiden, wenn er Produkte bzw. Unternehmen unterscheiden kann. Der Produzent kann nur dann Qualitätsstandards einhalten und Werbebotschaften übermitteln, wenn er sicher sein kann, dass sein Kennzeichen bzw. das Kennzeichen seines Produktes nicht durch entsprechende Kennzeichen anderer Hersteller verwässert oder gar bedroht wird.
Es lassen sich drei verschiedene Arten von Kennzeichen unterscheiden: **Die Marke** (§ 3, 4 MarkenG), **die geschäftliche Bezeichnung** (§ 5 MarkenG) sowie **die geographische Herkunftsangabe** (§ 126 ff. MarkenG).

Als **Marke** können alle Zeichen, insbesondere Wörter einschließlich Programmnamen, Abbildungen, Buchstaben, Zahlen, Hörzeichen, dreidimensionale Gestaltungen einschließlich der Form einer Ware oder ihrer Verpackung sowie sonstige Aufmachungen einschließlich Farben und Farbzusammenstellungen geschützt werden. Einzige Bedingung ist, dass diese geeignet sind, Waren oder Dienstleistungen eines Unternehmens von denjenigen anderer Unternehmen zu unterscheiden (§ 3 MarkenG). Inhaber einer Marke kann jede natürliche oder juristische Person sowie jede rechtsfähige Personengesellschaft sein. Der Markenschutz entsteht entweder durch die Eintragung eines Zeichens in das vom Deutschen Patentamt geführte Markenregister, durch Benutzung und Erwerb von Verkehrsgeltung oder (z. B. für Weltunternehmen) durch den Erwerb notorischer

Bekanntheit (§ 4 MarkenG). Liegt einer der drei Sachverhalte vor, so hat der Inhaber der Marke ein ausschließliches Recht, welches ihm erlaubt, Dritten zu untersagen, ohne seine Zustimmung ein mit der Marke identisches Zeichen für ihre Waren oder Dienstleistungen zu benutzen. Auch dürfen Dritte kein ähnliches Zeichen benutzen falls dadurch die Gefahr von Verwechslungen besteht (§ 14 MarkenG). Sollte es dennoch zum Kollisionsfall von Markenrechten kommen, so gilt der Prioritätsgrundsatz: Das prioritätsältere Recht hat demnach Vorrang vor dem prioritätsjüngeren Recht (§ 14 MarkenG). Hierbei stehen dem prioritätsälteren Recht Löschungsansprüche zu.
Bei eingetragenen Marken erstreckt sich der Markenschutz auf das gesamte Bundesgebiet. Ist eine Marke nicht eingetragen, so gilt der Markenschutz im räumlichen Bereich der Verkehrsgeltung. Bei Markenrechtsverletzungen hat der Markeninhaber das Recht, Dritte auf Unterlassung in Anspruch zu nehmen (§14 MarkenG). Sollte der Dritte fahrlässig oder gar vorsätzlich gehandelt haben, so steht dem Inhaber ein Schadensersatzanspruch in voller Höhe des durch die Verletzung entstandenen Schadens zu.
Bei Marken gilt es zu unterscheiden zwischen der sog. Individualmarke und der sog. Kollektivmarke. Bei der Individualmarke ist der Inhaber eine natürliche Person, eine juristische Person oder eine rechtsfähige Personengesellschaft. Bei der Kollektivmarke ist der Inhaber ein Verband, wobei die Marke die Waren der Verbandsmitglieder identifiziert.

Zu den **geschäftlichen Bezeichnungen** gehören die sog. Unternehmens-kennzeichen, welche nicht das Produkt, sondern das Unternehmen identifizieren. Ein Unternehmenskennzeichen entsteht nicht durch Eintragung, sondern durch Benutzung bzw. Erwerb von Bekanntheit (Verkehrsgeltung). Es lassen sich vier Arten von Unternehmenskennzeichen unterscheiden: Der Name (identifiziert Personen und Unternehmen), die Firma (identifiziert Kaufleute), die Besondere Geschäftsbezeichnung (z. B. „Hotel zur Post“) und das Geschäftsabzeichen (Kennzeichen, die nicht per se den Geschäftsbetrieb identifizieren, z. B. Aufmachung des Ladens, Kleidung der Angestellten). Der Inhaber einer geschäftlichen Bezeichnung kann es einem Dritten untersagen, eine gleiche oder ähnliche geschäftliche Bezeichnung zu benutzen falls dadurch die Gefahr der Verwechslung, Verwässerung oder der Rufschädigung besteht (§ 15 MarkenG). Benutzt ein Dritter

die gleiche geschäftliche Bezeichnung, so kann der Rechteinhaber, den Dritten auf Unterlassung in Anspruch nehmen (§15 MarkenG). Sollte der Dritte fahrlässig oder gar vorsätzlich gehandelt haben, so steht dem Rechteinhaber ein Schadensersatzanspruch in voller Höhe des durch die Verletzung entstandenen Schadens zu.

Die **Geographische Herkunftsangabe** weist wie der Name sagt, auf die Herkunft einer Ware oder Dienstleistung hin (z. B. Champagner). Dieser Begriff ist nicht zu Verwechseln mit dem Begriff der Gattungsbezeichnung. Die Gattungsbezeichnung bezeichnet das Produkt an sich (z. B. Pils, Kölnisch Wasser), stellt somit den Gegenbegriff zur Geographischen Herkunftsangabe dar und wird rechtlich nicht geschützt. Wer im geschäftlichen Verkehr geographische Herkunftsangaben benutzt, kann vom Berechtigten nach § 13 Abs. 2 des Gesetzes gegen den unlauteren Wettbewerb (UWG) auf Unterlassung in Anspruch genommen werden. Handelt der Nichtberechtigte fahrlässig oder vorsätzlich, so ist er dem Berechtigten zum Ersatz des dadurch entstandenen Schadens verpflichtet.

## 3.2 Kennzeichenrecht und Internet-Domains:

Um sich als Privatperson oder Firma im Internet präsentieren zu können benötigt man eine Homepage, auf welche die anderen Internet-Nutzer zugreifen können. Voraussetzung für den Zugriff auf die Homepage ist eine Art Adresse, die sog. Domain. Diese Domain setzt sich in den meisten Fällen aus drei Bestandteilen zusammen: Dem Kürzel für das World Wide Web (www), dem eigentlichen Namen der Domain (der sog. Second Level Domain) und einer landesspezifischen Endung (genannt Top Level Domain). Es ergeben sich also Adressen der Form www.SecondLevelDomain.TopLevelDomain, z. B. „www.ulm.de“. Damit die Domain dem Nutzer zur Verfügung steht, muss sie bei der entsprechenden Vergabestelle registriert werden. In Deutschland ist die „DENIC e.G“ (Deutsches Network Information Center; www.denic.de) für die Domainvergabe zuständig. Seit Bestehen des oben erläuterten Vergabesystems kommt es regelmäßig zu Rechtsstreitigkeiten bzgl. des Domainnamens, z. B. dann, wenn sowohl eine Privatperson als auch eine Firma mit gleichem Namen ihren Anspruch auf die entsprechende Second Level

Domain anmelden. Um diesen Ärger zu vermeiden sollte zuerst beim Deutschen Patent- und Markenamt bzw. im Handelsregister nachgeforscht werden, ob dort Marken oder Firmen angemeldet bzw. eingetragen sind, deren Namen identisch mit dem der Wunsch-Domain sind. Auch bei starker Ähnlichkeit der Namen kann es durchaus zu Problemen kommen. Hilfeleistung bzgl. dieser Nachforschungen bieten z. B. Patent-Informationszentren des Amtes, kommerzielle Markenrecherchedienste oder auch Patent- bzw. Rechtsanwälte an. Weitere Informationen hierzu sind auf der Homepage des Deutschen Patent- und Markenamtes (www.dpma.de) erhältlich. In Zusammenhang mit der Registrierung der Domains kommt es immer wieder zu rechtswidrigen Handlungen, die als „Domaingrabbing“ bezeichnet werden. Domaingrabber reservieren sich meist eine Vielzahl von Domains ohne diese dann selbst zu nutzen. Ihr Ziel ist es vielmehr die Domains später mit Gewinn an interessierte Personen oder Unternehmen weiterzuverkaufen. Gegen diese Art von rechtswidrigen Handlungen gibt es allerdings einige gesetzliche Vorgehensmöglichkeiten:

Schutz nach dem Namensrecht:

Hier bietet § 12 BGB einen entsprechenden Schutz. Dieser besagt, dass der Berechtigte von dem anderen die Beseitigung der Beeinträchtigung verlangen kann, wenn das Recht zum Gebrauch eines Namens von dem anderen bestritten wird oder das Interesse des Berechtigten dadurch verletzt wird, dass ein anderer unbefugt den gleichen Namen gebraucht. Die Nutzung des Domainnamens stellt definitiv einen solchen Gebrauch i. S. d. § 12 BGB dar. Die Anwendung des § 12 BGB ist jedoch nur dann möglich, wenn der Domainname identisch mit dem betroffenen Firmennamen ist oder zumindest starke Ähnlichkeit aufweist. Es muss eindeutig zu einer Zuordnungsverwirrung gekommen sein, oder durch die Wahl des Domainnamens eine Verbindung zum Produkt bzw. der Dienstleistung eines Unternehmens suggeriert werden, die in Wahrheit nicht besteht. Beispielsweise konnte der Versandhändler "Neckermann" die Herausgabe der Domain ,,Neckermann.de" verlangen. Unter dieser Domain-Adresse hatte der bisherige Inhaber versucht eigene Versandkataloge zu vertreiben.

Schutz nach dem Markenrecht:

Gemäß § 15 MarkenG ist es Dritten untersagt, die geschäftliche Bezeichnung oder ein ähnliches Zeichen im geschäftlichen Verkehr unbefugt in einer Weise zu benutzen, die geeignet ist, Verwechslungen mit der geschützten Bezeichnung hervorzurufen. Ob der Markeninhaber einen Anspruch gegen den Dritten hat hängt vor allem davon ab, ob der Domainname überhaupt im geschäftlichen Verkehr genutzt wurde und ob eine Identität oder starke Ähnlichkeit zwischen dem Domainnamen und der geschützten Marke besteht. So sprach beispielsweise das LG München der überregionalen Zeitschrift "Freundin" einen Anspruch gegen den Inhaber der Domain "freundin.de" (unter diesem Domain-Namen betrieb eine Partnervermittlung ihre Homepage) zu.

Schutz durch das Gesetz gegen unlauteren Wettbewerb (UWG):

§ 1 UWG besagt, dass derjenige, der im geschäftlichen Verkehre zu Zwecken des Wettbewerbes Handlungen vornimmt, die gegen die guten Sitten verstoßen, auf Unterlassung und Schadensersatz in Anspruch genommen werden kann. Unbedingt notwendig ist hierzu allerdings das Vorliegen eines Wettbewerbsverhältnisses zwischen den beiden Parteien. Ein Wettbewerbsverhältnis wird charakterisiert durch die Wechselbeziehung zwischen mindestens zwei Firmen, wobei die Handlung des einen für die Förderung seiner Firma auf Kosten der anderen Firma geht. Die bloße gemeinsame Benutzung des Internets ist dafür allerdings nicht ausreichend.

Sofern sich der Domaingrabber aufgrund der o.g. Ansprüche nicht zur Freigabe der entsprechenden Domain-Adresse entscheidet, müssen mögliche Ansprüche gegen ihn mittels Klage durchgesetzt werden. Je nach Darstellungen kann dann Unterlassung der Domain-Nutzung oder auch Schadenersatz verlangt werden. Auf jeden Fall sollte schnellstmöglich bei der zuständigen Domain-Vergabestelle (DENIC) ein sog. „Dispute“ oder „Wait“ eingerichtet werden. Dieser Antrag bewirkt vor allem, dass der Domaingrabber die Domain nicht auf einen Dritten übertragen kann. Dadurch wird verhindert, dass sich der Domaingrabber einer Auseinandersetzung mit dem Antragsteller entzieht. Außerdem wird gewährleistet, dass der Antragsteller des Dispute nach der Freigabe der Domain automatisch neuer Inhaber wird. Weder die Einrichtung noch die Aufhebung des Dispute sind dabei kostenpflichtig.

Quellenangaben zu diesem Kapitel:

Bürgerliches Gesetzbuch

Deutsches Markengesetz

Gesetz gegen unlauteren Wettbewerb

e-Recht24.de: www.e-recht24.de/druck-1-5.html

Kanzlei Panke und Partner: www.pauke-eurojuris.de/rechtsgebiete/internet.html

Dr. Christopher Lieb: „Marken- und Kennzeichenrecht“ unter www.wlk-erlangen.de/aufsaetze/aufs_content/wlk_aufs092.html

# 4. Datenschutz im Internet:

## 4.1 Allgemeines:

Die Kommunikation über das Internet birgt viele Gefahren für den Identitätsschutz, die Anonymität des Nutzers und auch für die Vertraulichkeit des Kommunikationsinhalts. So können hinterlassene Datenspuren nicht nur Auskunft über die Beteiligten eines Kommunikationsprozesses geben, sondern auch über dessen Inhalt, z. B. welche Waren in einem Online-Shop bestellt wurden. Trotzdem werden bei vielen Unternehmen, die im World Wide Web ihre Waren und Dienstleistungen anbieten, die Rechtsvorschriften bzgl. des Datenschutzes gar nicht, oder nur teilweise beachtet. Ein möglicher Grund dafür ist, dass sich der Datenschutz zu einem beachtlichen Teil aus dem Persönlichkeitsschutz herleitet. Dies bedeutet, dass der Datenschutz dem Nutzer dient und nicht dem Unternehmen, welches die Daten erhebt. Aus wirtschaftlicher Sichtweise ist dieses Verhalten der Unternehmen mehr als unverständlich, hängt doch der Erfolg eines Unternehmens, das online personenbezogene Daten (persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person, z. B. Name, Anschrift, Geburtsdatum, email-Adresse) verarbeitet und nutzt, stark vom Vertrauen der Nutzer in die Seriosität des Anbieters und der Transparenz des Angebotes ab. Eine vollständige Umsetzung der Datenschutzvorschriften und –regelungen könnte durchaus als eine vertrauens-

bildende Maßnahme gegenüber dem Kunden gesehen werden und dem Unternehmen helfen, die ohnehin schon starke Unsicherheit der Nutzer auf diesem Gebiet abzubauen. Die Einhaltung der verschiedenen Datenschutzvorschriften bzw. –regelungen wäre also für beide Seiten überaus vorteilhaft. Allerdings unterliegen nicht nur personenbezogene Daten dem Datenschutz, sondern auch Daten, die scheinbar gar nicht personenbezogen sind, wie z. B. eine IP-Adresse oder die diversen Log-Files, die auf den Web-Servern jeden einzelnen Zugriff speichern. Auch gilt es zu beachten, dass personenbezogene Daten nicht nur manuell durch die Nutzereingabe erhoben werden. Oftmals werden diese Daten automatisch aus den Browsern ausgelesen und dann in sog. Cookies gespeichert. Diese Cookies werden dann auf den Rechnern der Nutzer oder auf Servern abgelegt und ausgewertet. Gesetzlich ausdrücklich erlaubt ist dagegen z. B. die Verarbeitung von Bestandsdaten, die zur Ausgestaltung oder Änderung von Verträgen zwischen dem Nutzer und dem Provider erforderlich sind (siehe § 5 TDDSG). Gemäß § 6 TDDSG dürfen des weiteren auch Daten erhoben werden, die dem Nutzer die Inanspruchnahme von Telediensten ermöglichen (sog. Nutzungsdaten). Auch die Erhebung von Abrechnungsdaten ist zulässig. In diesem Zusammenhang ist jedoch zu beachten, dass Nutzungsdaten unmittelbar nach dem Ende der jeweiligen Nutzung, Abrechnungsdaten spätestens 80 Tage nach Rechnungsversand gelöscht werden müssen. Im folgenden soll ein kurzer Überblick über die verschiedenen Gesetze, Verordnungen und Regelungen bzgl. des Datenschutzes gegeben werden und die sich daraus ergebenden Pflichten für die Dienstanbieter betrachtet werden.

Es gilt zunächst drei große Bereiche zu unterscheiden, die von unterschiedlichen Gesetzen und Verordnungen geprägt sind, und somit jeweils für den Datenschutz einer bestimmten Art von Internetdiensten zuständig sind.

Für den ersten Bereich sind das Telekommunikationsgesetz (TKG) und die Telekommunikationsdatenschutzverordnung (TDSV) maßgebend. Der Zweck des Telekommunikationsgesetzes ist es gemäß § 1 TKG, durch Regulierung im Bereich der Telekommunikation den Wettbewerb zu fördern und flächendeckend angemessene und ausreichende Dienstleistungen zu gewährleisten sowie eine Frequenzordnung festzulegen. Die TDSV regelt dagegen den Schutz personenbezogener Daten der an der Telekommunikation Beteiligten bei der Erhebung, Verarbeitung und Nutzung dieser Daten durch Unternehmen und Personen, die geschäftsmäßig Telekommunikationsdienste erbringen oder an deren

Erbringung mitwirken (§1 TDSV). Dieser Bereich regelt die datenschutzrechtlichen Aspekte der sog. Transportdienste**,** also derjenigen Dienste, die sich auf die reine Transportebene wie z.B. email-Transport oder Online-Telefonie spezialisiert haben.
Der zweite große Bereich ist durch das Teledienstegesetz (TDG) und durch das Teledienstedatenschutzgesetz (TDDSG) charakterisiert. Ziel des TDG ist es, einheitliche wirtschaftliche Rahmenbedingungen für die verschiedenen Nutzungsmöglichkeiten der elektronischen Informations- und Kommunikationsdienste zu schaffen (§ 1 TDG). Das TDDSG sorgt dem entsprechend für den Schutz personenbezogener Daten bei den sog. Telediensten. Diese sind durch die individuelle Nutzung des einzelnen Kunden gekennzeichnet. Hier wären beispielhaft das e-Banking oder die elektronische Abwicklung von Bestellungen in e-Shops zu nennen.
Für den dritten Bereich ist der Mediendienstestaatsvertrag (MDStV) maßgebend. Dessen Zweck ist es nach § 1 MDStV, in allen Ländern einheitliche Rahmenbedingungen für die verschiedenen Nutzungsmöglichkeiten der elektronischen Informations- und Kommunikationsdienste zu schaffen. Dadurch wird der Datenschutz der sog. Mediendienste geregelt. Mediendienste richten sich an eine Vielzahl von Nutzern. Im Mittelpunkt steht die redaktionelle Bearbeitung sowie die elektronische Verteilung. Als Beispiel könnte man sich hierzu Online-Auftritte von Zeitungen oder auch Unternehmenspräsentationen vorstellen.

## 4.2 Datenschutzrechtliche Pflichten der Unternehmen:

Aus den o. g. drei Bereichen und ihren relevanten Bestimmungen für die einzelnen Internetdienste ergeben sich für die Unternehmen eine Vielzahl von Pflichten. Hier werden nur die wichtigsten erläutert.
Unter der **Impressumspflicht** wird die Pflicht verstanden, den Anbieter einer Website im Impressum zu kennzeichnen. Dabei sollten folgende Angaben gemacht werden: Name und Anschrift, Vertretungsberechtigter, Tel-Nr. und / oder email-Adresse, Register bzw. Registernummer, Umsatzsteueridentifikationsnummer.
Für die Unternehmen besteht auch eine **Pflicht zur Unterrichtung des Nutzers bzgl. der Datenspeicherung**. Diese Unterrichtung muss zu Beginn des Nutzungsvorgangs erfolgen. Die Nutzer müssen darüber in Kenntnis gesetzt werden,

welche Arten personenbezogener Daten über sie erhoben, verarbeitet oder benutzt werden sollen. Des weiteren hat die Unterrichtung vollständig und verständlich zu sein, d.h. sie muss in ausreichender Schriftgröße, hinreichend auffällig und in deutscher Sprache auf der entsprechenden Website angebracht sein. Besonders wichtig ist hierbei, dass der Nutzer für die Erhebung, Verarbeitung und Nutzung seiner personenbezogenen Daten ganz ausdrücklich sein Einverständnis erklären muss. Diese Einwilligung bedarf grundsätzlich der Schriftform. Gemäß § 3 TDDSG kann jedoch die Einwilligung auch elektronisch erklärt werden, wenn der Diensteanbieter sicherstellt, dass sie nur durch eine eindeutige und bewusste Handlung des Nutzers erfolgen kann (beispielsweise durch Klick auf einen Button), sie nicht unerkennbar verändert werden kann, ihr Urheber erkannt werden kann, die Einwilligung protokolliert wird und der Inhalt der Einwilligung jederzeit vom Nutzer abgerufen werden kann. Weiterhin besteht eine **Auskunftspflicht gegenüber dem Nutzer bzgl. der über ihn gespeicherten Daten**.

Ausserdem besteht für Unternehmen, in denen mindestens 5 Mitarbeiter ständig mit der Datenverarbeitung beschäftigt sind, die Pflicht, einen **Datenschutzbeauftragten** zu stellen.

Die Einhaltung dieser eben genannten Pflichten wird durch die örtliche Datenschutzaufsicht überwacht. Diese ist berechtigt, jederzeit Überprüfungen durchzuführen, auch wenn keine konkrete Beschwerde vorliegt. Bei Nichtbefolgung dieser Pflichten drohen enorme Bußgelder. So sieht beispielweise § 43 BDSG ein Bußgeld von bis zu 250.000 € vor.

<u>Quellenangaben zu diesem Kapitel:</u>

Teledienstedatenschutzgesetz

Teledienstegesetz

Telekommunikationsgesetz

Telekommunikations-Datenschutzverordnung

Mediendienstestaatsvertrag

Bundesdatenschutzgesetz

Weinknecht, OJR, Ausgabe 2000/11 unter www.weinknecht.de/ojr.

## Literatur- und Quellenverzeichnis:

Prof. Dr. Wolfram Gass: Skript zur VL „Medienrecht“, Uni Ulm, 2003

Deutsches Urhebergesetz

Gesellschaft zur Verfolgung von Urheberrechtsverletzungen e.V.: www.gvu.de/deutsch/fr_recht.html

Kanzlei Panke und Partner: www.panke-eurojuris.de/rechtsgebiete/urheber.html

Bürgerliches Gesetzbuch

Deutsches Markengesetz

Gesetz gegen unlauteren Wettbewerb

e-Recht24.de: www.e-recht24.de/druck-1-5.html

Kanzlei Panke und Partner: www.pauke-eurojuris.de/rechtsgebiete/internet.html

Dr. Christopher Lieb: „Marken- und Kennzeichenrecht“ unter www.wlk-erlangen.de/aufsaetze/aufs_content/wlk_aufs092.html

Teledienstedatenschutzgesetz

Teledienstegesetz

Telekommunikationsgesetz

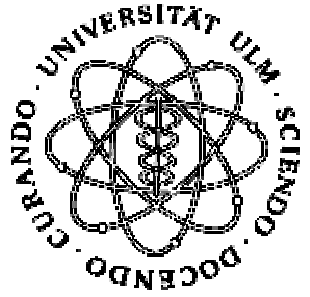
Telekommunikations-Datenschutzverordnung

Mediendienstestaatsvertrag

Bundesdatenschutzgesetz

Weinknecht, OJR, Ausgabe 2000/11 unter www.weinknecht.de/ojr.

# Seminararbeit im Schwerpunktfach Informatik
# SS 03




Thema : „Internetrecht : Online Auktionen“

Betreuer : Prof. Dr. Schweiggert, Abteilung SAI

Verfasser : Adam Kulas, Wirtschaftswissenschaften,
8. Semester, E-mail: kulas@mathematik.uni-ulm.de

Inhaltsverzeichnis

**1 Einleitung**

1.1 Motivation ..... 3

1.2 Internetdienste : Online Auktionen ..... 4

1.3 Folgerung ..... 4 - 5

**2 Gerichtsurteil zum Thema „Online Auktion“**

2.1 Tatbestand..... 5 - 6

2.2 Forderung des Klägers ..... 6

2.3 Argumentation des Beklagten ..... 6 - 7

2.4 Ausführungen des Gerichts, Urteil .....7 –12

**3 Vorsichtsmaßnahmen** ..... 12 -13

**Quellenangaben** ..... 13

# 1 . Einleitung

## 1.1 Motivation

Wenn man Äußerungen in den Medien zum Computerrecht aufmerksam genug verfolgt, so lässt sich feststellen , dass mit steigender Akzeptanz und kommerzieller wie privater Nutzung von Internetdiensten , gesetzliche Regelungen und Rechtsbestimmungen zunehmend erörtert werden.
Dass ein sensibilisiertes Bewusstsein für die rechtliche und persönliche Verantwortung bei der Verbreitung von Text-, Bild- oder Audio-Informationen oder bei persönlichen Meinungsäußerungen besteht, lässt sich durch Beobachtungen im Internet nicht feststellen.
Es ist aber unerlässlich zumindest einige Grundkenntnisse zu besitzen, wenn man nicht Gefahr laufen möchte, mit etwaigen Schadensersatzforderungen konfrontiert zu werden.
Die Annahme, es gäbe überhaupt kein gültiges Recht im Internet ist schlichtweg falsch. Auch wenn (noch) keine Spezialgesetze existieren, so finden doch die ansonsten bestehenden Regeln Anwendung, wie z.B. die des Zivil- und des Strafrechts.
Orte, an denen Informationen eingegeben und gelesen werden, unterliegen den normalen Regeln, ungeachtet dessen, ob es sich um einen virtuellen Raum handelt oder nicht. Somit ist das Internet weder staats- noch rechtsfrei.
Findet der Datenverkehr international statt, so ist es oft fraglich, welches Recht angewandt werden soll.
Zweck der Seminararbeit ist es ein Rechtsbewusstsein im Umgang mit dem Internet und den darin angebotenen Diensten, insbesondere Verkaufsplattformen, zu schaffen, um mögliche Rechtsstreitigkeiten vermeiden zu können.

## 1.2 Internetdienste : Online Auktionen

Im Februar zählten 32,5 Millionen Deutsche laut At-Facts-Umfrage zu den Internet-Nutzern. Knapp 40 Prozent der Onliner sind dabei täglich im Netz. Als reichweitenträkste Angebote ermittelte die Erhebung im Februar Ebay (58,2 Prozent der Nutzer), Google (57,7 Prozent) und das RTL-Netzwerk. (46,3 Prozent). (Newsletter eMarket vom 17.03.03, Ruth Stüber).
Das bedeutet, dass in etwa ein Viertel der Bevölkerung der Bundesrepublik zumindest die Seiten von Ebay besucht, oder als Käufer bzw. Verkäufer agiert.
Die steigende Tendenz nationaler und internationaler Internetnutzer von Online Auktionen Gebrauch zu machen spiegelt sich im Erfolg von Ebay wider, welches hier als Paradebeispiel für virtuelle Auktionshäuser angeführt werden soll, da es sich wohl um das beliebteste handelt.
Das Unternehmen wies für das abgelaufenen Quartal einen Reingewinn von 104,2 Mio. Dollar oder 32 Cent je Aktie nach 47,6 Mio. Dollar oder 17 Cent je Aktie im Vorjahr aus. Der Umsatz stieg den Angaben zufolge auf 476,5 Mio. Dollar von 245,1 Mio. Dollar im Jahr zuvor (n-tv , 22. April 2003).

## 1.3 Folgerung

Aufgrund der zunehmenden Beliebtheit von Verkaufsplattformen gebrauch zu machen, liegt es nahe, sich mit den damit resultierenden Pflichten und Rechten auseinander zu setzen.
Im Rahmen seiner privaten Umgebung fühlt sich der Mensch als Internetnutzer geschützt und anonymisiert. Er ist sich häufig nicht im Klaren darüber, dass ein Mausklick schwere rechtliche Folgen nach sich ziehen kann. Per Mausklick können, z.B. problemlos rechtsgültige Verträge geschlossen werden.

Die nachfolgende Darstellung eines Gerichtsfalles soll dazu beitragen eine gewisse rechtliche Verantwortung beim Umgang mit dem, nicht rechtsfreien, Internet zu schaffen und die Folgen unüberlegten Handelns darzulegen.

## 2 . Gerichtsurteil zum Thema : „Online Auktion“

(Bemerkung : Die nachfolgende Beschreibung des Urteils ist nicht vollständig und enthält Aspekte, die aus Sicht des Autors wichtig erscheinen.)

### 2.1 Der Tatbestand

Die Parteien streiten darüber, ob sie im Internet einen wirksamen Vertrag miteinander geschlossen haben.
Die Firma X in Hamburg verkauft über das Internet eigene Gegenstände gegen Höchstgebot, vermittelt auf diesem Wege Vertragsabschlüsse mit anderen Anbietern und gibt unter der Bezeichnung "X  private Auktionen" auch Dritten die Möglichkeit, eigene Verkaufsveranstaltungen durchzuführen.
Eine Teilnahme ist Internetbenutzern nur nach Anerkennung der Allgemeinen Geschäftsbedingungen für X - Verkaufsveranstaltungen (AGB) möglich. Auf diese AGB wird bereits auf der Homepage von X hingewiesen. Die Teilnehmer können durch zweimaliges Anklicken den Text der AGB in druckgerechter Form abrufen .
Der Beklagte hat unter "X private Auktionen" eine eigene Verkaufsveranstaltung durchgeführt und als Autohändler und anbietender Teilnehmer einen Neuwagen mit der Beschreibung: Passat Variant TDI 110 PS Neuwagen Trendline, Satinsilber metallic, innen schwarz, Edelholzausstattung, Klimaautomatic, Technik- und Winterpaket, Radiovorbereitung, Nebelscheinwerfer etc, ohne Zulassung, 0 Km, bei einem Startpreis von 10,-- DM ohne Angabe eines Mindestpreises vom 22.7.1999, 21.33 Uhr, bis zum 27.7.1999, 21.33 Uhr, angeboten.

Ein Neufahrzeug mit diesen Ausstattungsmerkmalen hat im Autohandel einen Listenpreis von ca. 57.000,-- DM.( Landgericht Münster, Aktenzeichen : 4 O 424/99 ,Entscheidung vom 21. Januar 2000).
Innerhalb der Bietzeit hat der Kläger als neunhundertdreiundsechzigster und letzter Bieter online ein Angebot über 26.350,-- DM abgegeben und von X eine Bestätigungs-eMail über seinen Neuerwerb erhalten.

2.2 Forderungen des Klägers

Der Kläger fordert die Herausgabe des Fahrzeugs zum Preis von 26.350 DM. Er beruft sich dabei auf die Allgemeinen Geschäftsbedingungen von X.
Auszug der AGB von X :
§ 5 Annahme eines Vertragsangebotes:
[...] (1) Der Vertrag über einen angebotenen Gegenstand kommt ohne Erklärung gegenüber dem Teilnehmer, der das Vertragsangebot abgegeben hat (nachfolgend auch "Antragender" genannt), bereits durch Annahme des Vertragsangebotes zustande. Der Antragende verzichtet auf eine Annahmeerklärung, § 151 Satz 1 BGB.

Das bedeutet, dass mit der Abgabe des höchsten Gebotes ein Vertrag mit dem Anbieter der zu ersteigernden Ware zustande kommt.

2.3 Argumentation des Beklagten

Im vorprozessualen Anwaltswechsel verweigerte der Beklagte, also der Anbieter des Fahrzeugs, die Herausgabe des Fahrzeugs zum Preis von 26.350 DM und berief sich dabei auf den §119 BGB [Anfechtbarkeit wegen Irrtums]. Dieser Paragraph besagt, dass „wer bei der Abgabe einer Willenserklärung (hier : Angebot zum ersteigern eines Autos)

über deren Inhalt im Irrtume war oder eine Erklärung dieses Inhalts überhaupt nicht abgeben wollte, die Erklärung anfechten kann, wenn anzunehmen ist, dass er sie bei Kenntnis der Sachlage und bei verständiger Würdigung des Falles nicht abgegeben haben würde.“ [...]

Der Beklagte meint, die AGB seien deswegen nicht anzuwenden, weil deren Kenntnisnahme durch die Teilnehmer an den Verkaufsveranstaltungen von X nur nach zahlreichen Schritten möglich sei.

Die Darstellung des Fahrzeuges im Internet sei kein Angebot im Sinne §145 [Bindung an den Antrag] : „Wer einem anderen die Schließung eines Vertrags anträgt, ist an den Antrag gebunden, es sei denn, dass er die Gebundenheit ausgeschlossen hat.“
Vielmehr sei es nur als Aufforderung zur Abgabe eines Angebots anzusehen (invitatio ad offerendum). Deswegen sei auch kein gültiger Vertrag zustande gekommen.

Weiter hätte sich der Beklagte irrtümlicherweise bei der Festsetzung des Startpreises geirrt. Anstelle von 10 DM , hätte er 10.000 DM tippen wollen. Eine spätere Berichtigung sei nicht möglich gewesen.
Auf die Festssetzung eines Mindestpreises hatte er laut Empfehlung von X verzichtet.

2.4 Ausführungen des Gerichts, Urteil

Das Landgericht Münster war der Auffassung, grundsätzliche Bedenken gegen die Zulässigkeit derartiger Online-Verkaufsaktionen unter Einbezug des Gewerberechts seien nicht angebracht. Das Fehlen einer gewerbeaufsichtlichen Genehmigung würde die privatrechtlich geschlossenen Rechtsgeschäfte nicht tangieren. §134[Gesetzliches Verbot] besagt: „Ein Rechtsgeschäft, das gegen ein gesetzliches Verbot verstößt, ist nichtig, wenn sich nicht aus dem Gesetz ein anderes ergibt.“

Da die privat geschlossenen Verträge gegen kein Gesetzt verstoßen, sind sie gültig .

Online abgegebene Erklärungen und auf diese Weise abgeschlossene Verträge seien nach dem BGB zu beurteilen. Für einen Vertragschluss bedürfe es zweier übereinstimmender Willenserklärungen , eines Angebotes und dessen Annahme. Diese Erklärungen könnten auch per Mausklick abgegeben werden. Die Präsentation des Autos sei , unter Berücksichtigung des BGB, aber noch kein Antrag auf Abschluss eines Vertrages, sondern eine Aufforderung zur Angebotsabgabe.
Aus Sicht des Gerichts sei zusätzlich die Endgültigkeit des rechtlichen Bindungswillens entscheidend.
Des Weiteren seien die AGB des Veranstalters heranzuziehen.
Mit der Beschreibung des Kaufgegenstandes durch den Beklagten entsprechend §3 AGB hat der Beklagte lediglich Interessenten aufgefordert Angebote anzugeben :

§3 Beschreibung des Kaufgegenstandes, Verkaufsangebot bei private Auktionen
(1) X ermöglicht es Teilnehmern, im Eigentum des jeweiligen Teilnehmers stehende Gegenstände, die im Rahmen von private Auktionen verkauft werden sollen, auf Angebotsseiten öffentlich zu präsentieren.

Aus der Sicht des Klägers könne deshalb die Präsentation nur als Aufforderung und nicht als bindendes Vertragsangebot gewertet werden. Daran ändere auch §5 (4) nichts :

(1) Der Vertrag über einen angebotenen Gegenstand kommt ohne Erklärung gegenüber dem Teilnehmer, der das Vertragsangebot abgegeben hat (nachfolgend auch "Antragender" genannt), bereits durch Annahme des Vertragsangebotes zustande. Der Antragende verzichtet auf eine Annahmeerklärung, § 151 Satz 1 BGB.

Da es sich seitens des Klägers mit dem Höchstgebot von 26.350 DM nur um ein Vertragsangebot handle, welches der Annahme seitens des Beklagten bedürfe, sei es zu keinem Vertragsabschluss gekommen. Insbesondere deshalb nicht, weil der Beklagte

weder ausdrücklich noch konkludent (durch eine bestimmte Handlung) eine Annahmeerklärung auf das Angebot abgegeben hat. Nach den AGB könne auch von keiner Bevollmächtigung von X seitens des Beklagten ausgegangen werden.

Die von X am 27.01.1999 an den Kläger gerichtete E-Mail stelle keine Annahmeerklärung dar, obwohl in ihr von einem Zustandekommen des Vertrages die Rede sei, denn das Schriftstück enthalte nicht die für einen Vertragsabschluss benötigte Annahmeerklärung des Beklagten:
"Gesendet: Dienstag, 27. Juli1999 21:54 Betreff: Herzlichen Glückwunsch! (Ihr Zuschlag bei der Auktion Nr., 174124). Hallo, herzlichen Glückwunsch, Ihr letztes Gebot war unschlagbar! Bei „X private Auktionen“ haben Sie für 26.350,00 DM den Zuschlag bei der Auktion von „X Automobile“ mit dem Titel VW Passat Variant TDI 110 PS - Neuwagen (Auktions-Nr 174124) erhalten. Bitte setzen Sie sich mit „X Automobile“ in Verbindung, um Versand und Bezahlung schnell und einfach zu regeln: Benutzername: Name: E-Mail: Telefon."

Eine Vertragsannahme könne auch nicht durch §5 (4) der AGB angenommen werden :
„§5 (4) Bei private Auktionen erklärt der anbietende Teilnehmer bereits mit der Freischaltung seiner Angebotsseite gemäß § 3 Abs. 5 die Annahme des höchsten unter Berücksichtigung von § 4 Abs. 4 und 5 wirksam abgegebenen Kaufangebotes.“
Dieser Erklärung sei nicht der Erklärungswert beizumessen, dass sich der Beklagte mit einem Verkaufspreis von 26.350 DM einverstanden erklärt hat.

Es sei der tatsächliche Wille des Beklagten zu erforschen laut §133 BGB [Auslegung einer Willenserklärung] : „Bei der Auslegung einer Willenserklärung ist der wirkliche Wille zu erforschen und nicht an dem buchstäblichen Sinne des Ausdrucks zu haften.“ Somit sei es fraglich, ob der Beklagte jedes innerhalb des Angebotszeitraumes abgegebene Vertragsangebot von vornherein annehmen wollte.

Einer Auslegung der Willenserklärung des Beklagten bedürfe es hierbei deshalb, da §5 (4) AGB abstrakt sei, denn er enthalte keine Angaben über einen Preis, die nach §433[Vertragliche Hauptpflichten] BGB nötig seien.

Die Annahmeerklärung des Beklagten müsse vom Kläger empfangen und nach Treu und Glauben unter Berücksichtigung der Verkehrssitte ausgelegt werden. Dabei sei auch der mit dem Rechtsgeschäft verfolgte Zweck und die bestehende Interessenslage zu berücksichtigen.
Die oben genannten Gründe führen also nicht dazu, dass der Beklagte das Angebot des Klägers angenommen hat. Darüber hinaus konnte der Kläger nicht davon ausgehen, dass es sich seitens des Beklagten z.B. um eine Werbeveranstaltung handelte, bei der mögliche Vermögenseinbußen seitens des Klägers in Kauf genommen werden würden. Er hätte allerhöchstens mit einen so genannten „Hauspreis“ rechnen dürfen.
Das fehlen der Eingabe eines Mindestpreises dürfe nicht so verstanden werden, dass sich der Beklagte mit dem Zustandekommen eines Preises in jeglicher Höhe einverstanden erklären würde, zumal der Anbieter X von der Eingabe eines Mindestpreises abrät, um mehr Interessenten anzulocken.

Es sei weiter zu berücksichtigen, dass der Anbieter nach der erstmaligen Einrichtung der Angebotsseite nur noch die Beschreibung der angebotenen Gegenstände ergänzen könne, während es X nach §6 AGB (...) möglich sei die Veranstaltung abzubrechen, zu verkürzen oder zu verlängern. Somit würde das höchste Gebot durch das Ende des Angebotszeitraumes bestimmt und nicht wie bei einer Versteigerung im Sinne der Gewerbeordnung durch die Preisvorstellungen der Interessenten und den Verkehrswert der Ware. Der Anbieter könne den Preis nur durch Dritte, die in seinem Sinne den Preis in die Höhe treiben , steuern.
Durch die Umstände, unter denen bei „X – private Auktionen“ Anbieter ihre Ware präsentieren könnten, handele es sich eher um ein Glücksspiel. Das liege zum Einen an

der Begrenzung der Bietschritte auf max. 50 DM , zum Anderen am Verhalten erfahrener Bieter, die erst zum Ende der Veranstaltung anfangen ihre Gebote abzugeben. Ein „Ausbieten“ bis zum Höchstpreis sei aus diesem Grund nicht möglich, erst recht nicht wenn man bedenke, dass es einer Vielzahl von Geboten erfordere, um in einen realistischen Preisbereich zu kommen.

Zusammenfassend könne die vorweggenommene Annahmeerklärung des Beklagten redlicherweise nur so verstanden werden, dass er das Fahrzeug nicht unter seinen eigenen Kosten verkaufen wollte.

Zuletzt sei auch keine Verkehrssitte ersichtlich, die besagt, dass das zuletzt abgegebene höchste Gebot von den anbietenden Teilnehmer, hier dem Beklagten, in jedem Fall angenommen werden müsste. Bisher sei es der erste Fall dieser Art, der gerichtlich ausgetragen würde. Allerdings könnten im Normalfall über „X – private Auktionen“ rechtsgültige Verträge zu Stande kommen, die keiner ausdrücklichen Bestätigung, wie in diesem Fall, bedürften. Diese Verträge würden durch eine stillschweigende Übereinkunft bestätigt, die sich im Bezahlen des zuletzt abgegebenen Gebotes und dem Empfangen der Ware widerspiegelten. Das sei aber im vorliegendem Fall nicht passiert, sondern ausdrücklich vom Beklagten verweigert worden.

Falls man jedoch in „X – private Auktionen“ eine Versteigerung im gewerberechtlichen Sinne sehen wolle - was das Landgericht Münster nicht tat – so müsste der Fall nochmals geprüft werden.

Aus dem oben beschriebenen Gründen wurde die Klage abgewiesen, der Beklagte musste das Auto nicht verkaufen.
Das Verfahren ging in die nächste Instanz. Das Oberlandesgericht Hamm entschied in dieser Sache , dass der Student das Auto zu dem zuletzt abgegebenen höchsten Gebot

verkaufen müsse. Die Ausführungen des Oberlandesgerichtes richteten sich nach den vom Landgericht Münster erörterten Sachverhalten, nur wurden diese grundlegend anders gewichtet und gewertet.

## 3 Vorsichtsmaßnahmen, Hinweise

- Bedenken Sie immer, dass Sie sich im Internet in keinem rechtsfreien Raum bewegen. Die Entscheidungen, die Sie treffen haben Rechtsgültigkeit und können zu ernsthaften rechtlichen Folgen führen.
- Bevor Sie Verträge über das Internet abschließen, lesen Sie diese sorgfältig bzw. beachten Sie die Allgemeinen Geschäftsbedingungen des Anbieters. Im Zweifellsfall senden Sie eine Anfrage mit Bitte um Erklärung, oder befragen sie einen Juristen.
- Bei Internetauktionen gilt insbesondere :
    - Falls Sie etwas wertvolles versteigern möchten vergewissern Sie sich, dass Sie einen Mindestpreis angegeben haben, auch wenn die Betreiber der Plattform über die das Geschäft abgewickelt wird etwas Anderes raten.
    - Geben Sie als Käufer nur dann Gebote ab, falls Sie ein tatsächliches Interesse an der Ware haben. Die Zusage per Mausklick ist rechtsgültig.
    - Wählen sie keine trivialen Passwörter und halten Sie ihre Passwörter geheim. Es besteht Gefahr von Missbrauch.
    - Haben Sie eine Ware ersteigert, so versuchen Sie nach dem ersten Kontakt die vom Verkäufer angegebenen Daten selbst zu überprüfen. Die Internetbetreiber sichern sich mit ihren Allgemeinen Geschäftsbedingungen so ab, dass Sie bei Betrug keine rechtliche Grundlage gegen diese haben.

- Verträge, die über das Internet geschlossen werden regelt, wie im normalen Geschäftsleben, das BGB.

Quellenangaben :


- 1.2 Internetdienste : Online – Auktionen :
  Newsletter eMarket vom 17.03.03, Ruth Stüber
  n-tv , 22. April 2003
- 2 Gerichtsurteil zum Thema : „Online Auktion“ :
  www.stroemer.de : Urteil des Landgerichts Münster,
  Aktenzeichen : 4 O 424/99
  BGB

Universität Ulm - SAI

# Seminararbeit

zum Thema

# Firewalls

bearbeitet von: Frank Weidlich

betreut von: Prof. Dr. Franz Schweiggert

und Dr. Matthias Grabert

Ulm, Juli 2003

# Inhaltsverzeichnis


**1 Einleitung 1**

1.1 Was ist eine Firewall 1

**2 Firewall-Typen 2**

2.1 Packet Filtering Firewall 2

2.2 Proxys bzw. Application Gateway Firewall 3

2.3 Andere Firewall Systeme 4

2.3.1 Stateful Packet Filter Firewall (SPF) 4

2.3.2 Stateful Inspection Filter 5

**3 Firewalls im praktischen Einsatz 5**

3.1 Ausschliesslicher Einsatz von Paket Filtern 5

3.2 Ausschliesslicher Einsatz von Proxys 6

3.3 Kombination von Proxy und Application Gateway 7

3.4 De-Militarized Zone (DMZ) 7

**4 Linux Firewall 8**

4.1 Allgemeines 8

4.2 Grundsätzlicher Aufbau 9

4.2.1 Tabellen 10

4.2.2 Chains 11

4.2.3 Tabellen-Chain Struktur 11

4.2.4 Stateful Firewalling 12

**5 Grenzen von Firewalls 13**

**6 Andere Sicherheitssysteme 13**

**7 Trends 14**

7.1 Firewalls basierend auf künstlicher Intelligenz 14

7.2 Post-Firewall Ära 14

**8 Fazit 15**

**Anhang Literatur- & Abbildungsverzeichnis 16**

# 1. Einleitung

Jede Organisation, ob sich nun um eine Behörde, eine Schule oder ein privates Unternehmen handelt, verfügt über eine enorme Menge an Informationen, die wenn sie in falsche Hände geraten, schwerwiegende Folgen für die Betroffenenen nach sich ziehen können. Ich denke hier vor allem an Informationen wie Krankheitsdaten, unternehmensinterne Daten (z.B. Höchstgebot bei UMTS-Lizenz-Versteigerung, u.a.), aber auch schon die Offenlegung von Gehaltsinformationen kann für viele Einzelne sehr unangenehm sein.
Gleichzeitig ist nicht nur die Weitergabe von Informationen für die Organisation ein enormes Problem, sondern die Zerstörung oder Modifikationen von Daten ist als mindestens genauso problematisch anzusehen. Ganz zu schweigen von den Auswirkungen, wenn ein Angreifer die Kontrolle über das ganze System bekommt.
Es besteht also hoffentlich kein Zweifel, dass der IT-Sicherheit eine elementare Wichtigkeit eingeräumt werden muss. Ich möchte nun in meinem Seminar auf das Konzept/Produkt/Einrichtung "Firewall" näher eingehen, um den Leser einen grundlegenden Einblick darüber zu geben.

## 1.1 Was ist eine Firewall?

Eine Firewall ist eine Einrichtung, die auf Netzwerkebene als Zugriffssteuerungs-mechanismus für ein oder mehrere bestimmte Netzwerke eingesetzt wird. In den meisten Fällen dienen Firewalls dazu, Unbefugten den Zugriff auf ein internes Netzwerk zu verweigern. Firewalls werden aber auch dazu verwendet, um innerhalb von LANs sichere Subnetze (z.B. für eine Entwicklungsabteilung) zu schaffen. Meistens sind Firewalls eigenständige Computer, Router oder sonstige Hardwaregeräte mit z.T. spezieller Software. Firewalls sind als Kontrollpunkte für das Netzwerk gedacht. Wenn Verbindungsanfragen gestellt werden, dann werden diese von der Firewall bearbeitet. Basierend auf einem vordefinierten Satz von Regeln (auch Richtlinien genannt) wird festgelegt, ob der Datenaustausch zulässig ist oder nicht.
(Vgl. anonymous, S.244f)

# 2. Firewall-Typen

## 2.1 Packet Filtering Firewall

Packet Filtering Firewalls (bzw. Paket Filter) sind oft Router, die über Funktionen zur Paketfilterung verfügen. Der Router bekommt Pakete von einem Netzwerk und leitet sie an ein anderes Netzwerk weiter. Zuvor vergleicht er die Pakete mit den vom Admininstrator definierten Regeln. Abhängig vom TCP- und IP-Header der Pakete und von den Regeln kann die Firewall das Paket ablehnen, es weiterleiten oder eine Nachricht zum Ursprung zurücksenden. Ein Ereignis, das gegen Filterregeln verstößt, wird protokolliert.

Kontrollmöglichkeiten der Firewall:

- Auf der Netzzugangsebene (im Intranet) werden Quell-, Zieladressen und der verwendete Protokolltyp kontrolliert.
- Auf der Netzwerkebene werden beim IP-Protokoll die Quell- und Zieladresse, sowie das Optionsfeld und Flags überprüft, außerdem die ICMP-Kommandos und die physikalische MAC-Adresse.
- Auf der Transportebene werden bei TCP und UDP die Portnummern kontrolliert, bei TCP zusätzlich die Richtung des Verbindungsaufbaus.

(Vgl. Pohlmann, S. 119ff)

Grundsätzlich können die Filterregeln generell entweder als Gebotsregeln oder als Verbotsregeln ausgelegt werden:

- Bei den Gebotsregeln ist alles, was nicht explizit erlaubt ist, verboten.
- Bei den Verbotsregeln ist alles, was nicht explizit verboten ist, erlaubt.

Dabei ist die erste Regel der zweiten Regel vorzuziehen , da diese wesentlich strenger in der Auslegung ist.

### Stärken von Paket Filtern:

Paketfilterung ist eine kostengünstige Technologie mit recht guter Performance. Ein Paketfilter ist heute auf fast allen Router-Produkten standardmäßig implementiert. Oft ist kein zusätzlicher Administrations- und Konfigurationsaufwand notwendig. Paket Filter sind leicht erweiterbar, wenn neue Dienste oder Protokolle transportiert werden müssen (hinzufügen neuer Regeln reicht im Normalfall).

### Schwächen von Paket Filtern:

Paketfilterregeln sind für den Durchschnittsbenutzer oft recht verwirrend. Bei großen Netzen können Filterregeln sehr umfangreich und schwer nachvollziehbar werden. Protokollmeldungen enthalten oft keine Informationen über den Inhalt der übertragenen und verworfenen Pakete. Einige Protokolle sind für Paket Filter ungeeignet, da variable Portnummern verwendet werden. Des Weiteren besitzen Portnummern und IP-Adressen eine unzureichende Integrität der, da diese leicht gefälscht werden können (IP-Spoofing). Ein weiterer Nachteil ist die fehlende Benutzerauthentifizierung und die fehlende Kontrolle der Inhalte der Datagramme.

## 2.2 Proxys bzw. Application Gateways

Ein Application Gateway trennt das sichere und das unsichere Netzwerk physisch, indem es zwei Netzwerkanschlüsse hat (einen Anschluss für das sichere und einen für das unsichere Netzwerk). Die logische Trennung zwischen dem sicheren und dem unsicheren Netzwerk wird dadurch realisiert, dass Benutzer im sicheren Netzwerk Kommunikationsanfragen und Daten an den Application Gateway richten, der Application Gateway die Kommunikationsanfragen und Daten analysiert und, wenn es die Sicherheitspolitik der Unternehmung erlaubt, diese an die eigentlichen Empfänger im unsicheren Netzwerk - stellvertretend für die Benutzer im sicheren Netzwerk - weiterleitet. Andersherum nimmt der Application Gateway auch Kommunikationsanfragen und Daten aus dem unsicheren Netzwerk entgegen, analysiert sie und leitet diese, unter Beachtung der Sicherheitspolitik der Unternehmung, an die eigentlichen Empfänger im sicheren Netzwerk, stellvertretend für die Kommunikationspartner im unsicheren Netzwerk, weiter. (Vgl. Pohlmann S. 146 f.; Hegering S. 197 ff.)

Für jeden Dienst ist ein spezifisches Proxyprogramm auf dem Proxy-Server erforderlich (z.B. FTP). Im Gegensatz zu Paket Filtern ist eine Nutzdatenanalyse möglich, das heißt, Daten können analysiert und z.B. nach bestimmten Schlüsselwörtern durchsucht werden (z.B. E-Mail, HTML-Seiten). Einige HTTP-Proxys bieten sogar die Möglichkeit, alle Zeilen innerhalb einer Seite, die zu Java-Applets gehören, zu löschen. Des weiteren können bestimmte Dienstmerkmale eingeschränkt werden. Meist ist eine Cache Funktionalität für Webseiten verfügbar. Im Gegensatz zu Paket-Filtern wird mit Application Gateways eine verlässliche Trennung zwischen unterschiedlich vertrauenswürdigen Netzwerk-

segmenten erreicht. Manipulierte Pakete werden zuverlässig abgefangen und entsprechend protokolliert.

### Stärken von Proxys / Application Gateways:

- Bieten ein hohes Maß an Sicherheit
- Sehr umfangreiche Protokollierung möglich
- Authentisierung des Benutzers
- Granularität auf Dienstebene
- Dienste können benutzerabhängig erlaubt werden
- Verbindung zwischen dem zu schützenden Netz und dem Internet wird durch den Application Gateway völlig entkoppelt

### Schwächen von Proxys / Application Gateways:

- Höherer Rechenaufwand nötig
- Wenig skalierbar
- Bei neuentwicklten Protokollen kann es (vorläufig zumindest) noch keinen entsprechenden Proxy geben

## 2.3 Andere Firewall Systeme:

### 2.3.1 Stateful Packet Filter Firewall (SPF)

Diese Firewalls basieren grundsätzlich auf dem Konzept der Paketfilterung, erweitern dieses aber noch um einige Aspekte. Auf diesem Modell beruhende Firewalls überwachen die Sitzungen und Verbindungen mit Hilfe interner Zustandstabellen und können dementsprechend sofort reagieren. Aus diesem Grund sind SPF-basierte Produkte flexibler als reine Paketfilterlösungen. Außerdem schützen die meisten SPF-basierten Firewalls auch vor bestimmten DoS-Angriffen[1] und stellen für SMTP-basierte Maildienste und andere sicherheitsrelevante Funktionen erweiterte Sicherheitsvorkehrungen bereit. (Vgl. anonymous, S. 251)



[1]Denial of Service -Angriffe: Nichtverfügbarkeit von Diensten

Mit SPF-basierten Firewalls kann man beispielsweise standardmäßig alle Ports über 1024 schließen. Diese Ports werden dann nur bei Bedarf geöffnet. (siehe auch Kap 4. Linux-Firewall)

### 2.3.2 Stateful Inspection Firewall

Stateful Inspection Firewalls kombinieren die Vorteile der Paketfilter- sowie der Proxy-Firewalls. Das Client/Server-Modell wird hier im Gegensatz zur Application Gateway-Firewall nicht durchbrochen, d.h. es schaltet sich kein Proxy zwischen den Client und den Server. Zuerst agiert diese Firewall vergleichbar mit einem Paket Filter. Dann werden zustandsbezogene Informationen aus den Anwendungsschichten extrahiert und in dynamischen Tabellen aufbewahrt, um damit nachfolgende Verbingungsaufbauversuche bewerten zu können. Hierbei werden nicht nur die Ports und Adressen untersucht, sondern auch der Datenstrom. Stateful Inspection Firewalls liegen in ihrer Performanz zwischen den  Paketfilter- und Proxylösungen. Sie sind eine Erweiterung der SPF-Firewalls (diese können den Datenstrom nicht untersuchen).

# 3. Firewalls im praktischen Einsatz

(Vgl. Pohlmann, S.189ff)

Allgemeine Grundsätze bei der Verwirklichung/ Implementation einer Firewall:

- Position der eingesetzten Geräte sollte möglichst weit außen sein
- Bei stark zu schützenden Netzen sollten Geräte redundant eingesetzt werden
    - Unterschiedliche Hersteller
    - Unterschiedliche Filterformate

Anzahl und Art der Firewallkomponenten sollte dem Sicherheitskonzept angepasst sein.

## 3.1 Ausschließlicher Einsatz von PaketFiltern

Dieses Konzept ermöglicht folgende Sicherheitsziele:

- Zugangskontrolle auf Netzzugangs- und Netzwerkebene (nur erlaubte logische Verbindungen können aufgebaut werden)

- Rechteverwaltung (Zugriff nur über erlaubte und definierte Protokolle und Ports)
- Protokollauswertung
- Alarmierung

Reine Packet Filter Lösungen bieten relativ wenig Schutz, um zu schützende Netze mit unsicheren Netzen zu koppeln (z.B. Gefahr durch Spoofing).

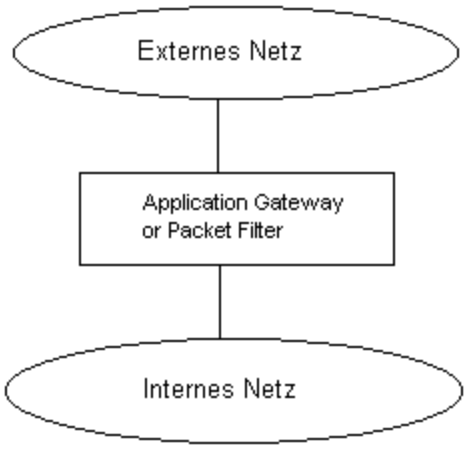



Abb. zu Abschnitt 3.1 und 3.2

## 3.2 Ausschließlicher Einsatz von Proxys

Hierdurch werden folgende Sicherheitsziele ermöglicht:

- Zugangskontrolle auf Netzwerkebene (s.o.)
- Rechteverwaltung (s.o. und nur solche Dienste sind möglich, für die ein Proxy eingerichtet wurde)
- Zugangskontrolle auf Benutzerebene (Benutzer können identifiziert und authentisiert werden)
- Kontrolle auf der Anwendungsebene (spezielle Anwendungsfilter für Kommandos, Anwendungsdaten und Dateien)
- Entkoppelung der Dienste (risikobehaftete Programme werden nur indirekt über Proxys dem unsicheren Netz zur Verfügung gestellt (z.B. Sendmail))
- Protokollauswertung und Beweissicherung (zusätzliche Beweissicherung von Handlungen einzelner Benutzer)
- Alarmierung
- Verbergen der internen Netzstruktur

Mit diesem Konzept können zwei Netze mit vergleichbarem Schutzniveau verbunden werden (z.B. Kooperation zweier gleichwertiger Organisationen).

## 3.3 Kombination von Paket Filter und Application Gateway

Um die Sicherheit zu erhöhen verbindet man hier einfach die beiden unterschiedlichen Firewalls. Der Weg aus dem unsicheren Netz in das zu schützende Netz läuft hier zuerst durch den Paket Filter und anschließend durch das Application Gateway. Es gibt keine Möglichkeit, das Application Gateway zu umgehen. Da beide Elemente mit unterschiedlichen Einbindungs- und Analysekonzepten arbeiten, werden sämtliche Sicherheitsziele erreicht und eine äußerst hohe Gesamtsicherheit realisiert. Sie verbinden die Vorteile der paketbasierten wie der proxybasierten Firewalls. Allerdings sind die Kosten höher und die Wartung und Einrichtung ist aufwendiger.

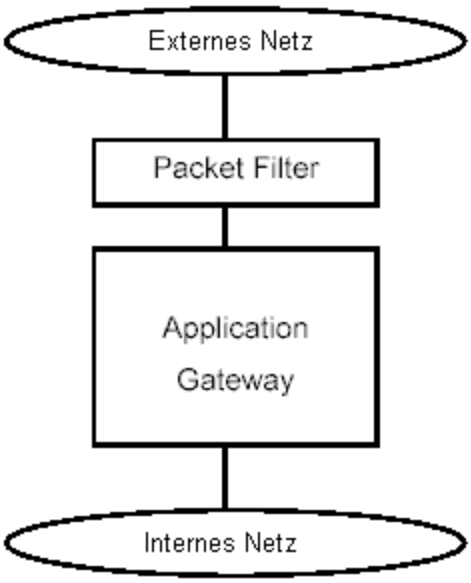



Abb. zu Abschnitt 3.3

## 3.4 De-Militarized Zone (DMZ)

Hier bedindet sich der Application Gateway im sog. Screened Subnet und wird aus den beiden angeschlossenen Netzen jeweils durch Paket Filter geschützt. Aus der Sicht der Paket Filter kommuniziert immer nur der Application Gateway mit dem jeweils

gekoppelten Netz, wodurch ihre Filterregeln einfach und überschaubar konfiguriert werden können. Durch die Anordnung muss ein Angreifer jetzt drei Barrieren überwinden, um in das zu schützende Netz zu gelangen. Es ist üblich, für die verschiedenen Elemente unterschiedliche Betriebssysteme einzusetzen, um evtl. Betriebssystemfehler oder Lücken auszugleichen. In diesem System können sämtliche Protokollebenen kontrolliert und analysiert werden. Es ist dafür geeignet, zu schützende Netze an sehr unsichere Netze, wie das Internet, zu koppeln. Im Screened Subnet ist es möglich, Internet Server z.B. zwischen den oberen Paket Filter und den Application Gateway oder direkt an den Gateway anzuschließen. Die Netze bleiben über den Application Gateway dabei immer vollständig entkoppelt.

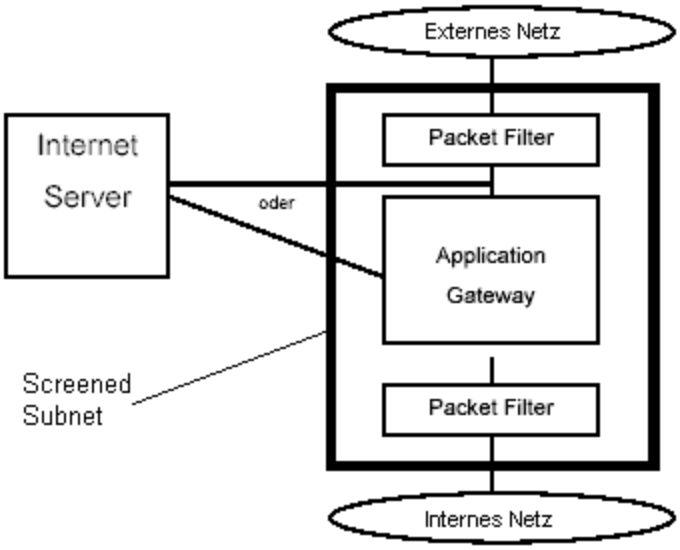



Abb. zu Abschnitt 3.4

Vorteile der Screened Subnet Architektur:

- Gute Skalierbarkeit
- Es müssen zwei Paket Filter überwunden werden (womöglich sogar unterschiedlicher Bauart)

Nachteile der Screened Subnet Architektur:

- Höhere Kosten
- Hoher Administrationsaufwand

# 4. Linux-Firewall

## 4.1 Allgemeines

Linux hat, im Gegensatz zu den meisten anderen Betriebssystemen, die Funktionalität der Firewall bereits im Betriebssystemkern (Kernel) fest eingebaut. Das macht die Firewall sowohl stabiler, als auch wesentlich schneller, als wenn sie in dem so genannten Userspace realisiert werden müsste.
Mit der Kernel 2.4 wurde auch die Firewall grundlegend überarbeitet, da die Vorgängerversion nicht mehr auf dem aktuellen Stand der Paketfilter-Technik war. Es wurde ein neues Firewall-Modul namens „Netfilter“ für den Kernel entwickelt. Die Entwickler bezeichneten ihr Werk als „Framework für die Manipulation von Datenpaketen“. Firewall-Entwickler bekommen mit ihm die Basisfunktionen zur Erweiterung von Netfilter an die Hand, ohne sich mit Kernelprogrammierung befassen zu müssen. Zwei grosse Vorteile von Netfilter sind:

- Das Framework selbst ist vollkommen protokoll-unabhängig und damit beliebig erweiterbar.
- An wesentlich weniger Stellen als bisher wird in den Kernel-Code eingegriffen, um die Firewall-Funktionalität unterzubringen, als dies vorher der Fall war. Damit ist der zurzeit vorliegende Code auch wesentlich besser wartbar als seine Vorgänger.

## 4.2 Grundsätzlicher Aufbau

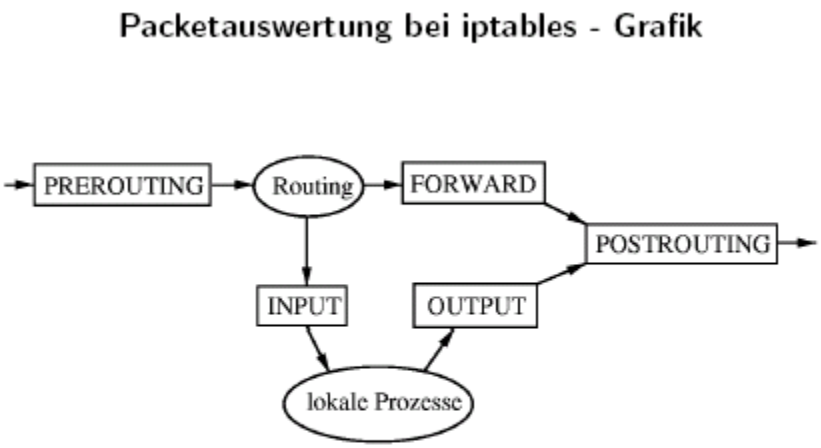
Die Firewall arbeitet mit einer Liste von Annahme- oder Ablehnungskriterien. Daraus entstehen so genannte Regeln, die genau bestimmen, ob ein Paket passieren darf oder nicht.

Nach folgenden Kriterien werden die Regeln aufgebaut:

- Netzwerkschnittstelle
- IP Adressen
- TCP/UDP Portnummern bzw. ICMP Nachrichtentypen
- SYN- und ACK-Flag

Netfilter ist also eine Serie von Regelketten (Chains) auf bestimmten Positionen in einem Protokoll-Stack wie das bereits beim Kernel 2.2.x der Fall war.

Kernelmodule können sich registrieren, um an irgendeiner dieser Regelketten zu lauschen. Wenn diese Netfilter Regelkette dann vom Networking-Code aufgerufen wird, hat an diesem Punkt jedes registrierte Modul die Möglichkeit, das Paket zu verändern. Der wesentlichste Unterschied zum Kernel 2.2.x bezüglich des Aufbaus der Firewall ist der, dass die Regelketten innerhalb des Protokoll-Stacks eine andere Position eingenommen haben und somit der Weg der Pakete durch die Firewall grundlegend verändert wurde. Durch die Veränderung des Weges, ändert sich somit auch der ganze Ablauf, welcher bei Netfilter wesentlich logischer aufgebaut ist als der Ablauf im Kernel 2.2.x.




Paketauswertung und -modifikation bei iptables

(entnommen aus Schreiber S. 6)

Doch nicht nur der Ablauf hat sich verändert, sondern auch an der Struktur der Firewall wurden diverse Veränderungen vorgenommen.

### 4.2.1 Tabellen

In der Netfilterarchitektur gibt es drei Tabellen. Diese Tabellen haben den Zweck, die verschiedenen Arten der Paketbehandlung auf Module verteilen zu können und nur die Module zu laden und damit auch nur die Tabellen mitzuführen, die im Augenblick, beziehungsweise für die gestellte Anforderung, benötigt werden.

Die drei Tabellen sind:

- Filter: Die Standard-Tabelle, die immer dann verwendet wird, wenn keine Tabelle explizit angegeben wird. Die Hauptaufgabe besteht aus dem Herausfiltern unerwünschter Datenpakete.
- Nat: Diese Tabelle ist für alle Arten von Adress-Umsetzungen oder Port-Forwarding verantwortlich.
- Mangle: In dieser Tabelle werden spezielle Änderungen an Paketen vorgenommen (z.B. Veränderung von Parametern wie time to live).

(Vgl. Thüring, S.11)

### 4.2.2 Chains

Jede Tabelle besteht aus mehreren, in seiner Funktion vordefinierten Regelketten. In den einzelnen Chains werden die einzelnen Regeln definiert, welche für die Pakete gelten sollen. Im Gegensatz zur alten Architektur, gibt es grundsätzlich fünf und nicht mehr drei, von der Architektur gegebene, Chains. Die fünf vorgegebenen Chains sind:

- INPUT: Hier landen alle Pakete, die an einen lokalen Prozess gerichtet sind.
- OUTPUT: Hier laufen alle Pakete durch, die von einem lokalen Prozess stammen.
- PREROUTING: Unmittelbar, bevor eine Routing-Entscheidung getroffen wird, müssen die Pakete hier durch.
- FORWARD: Behandelt alle zu routenden Pakete
- POSTROUTING: Alle Pakete, die geroutet werden, laufen nach dem Routing hier durch.

(Vgl. Thüring, S.11f)

Neben den fünf vordefinierten Chains, können auch beim Netfilter eigens definierte Chains erzeugt werden, in denen spezielle Regeln eingetragen werden können.

### 4.2.3 Tabellen-Chain Struktur

Da eine grosse Anzahl an Kombinationen zwischen Chains und Tabellen gebildet werden können, war es bei der Entwicklung wichtig eine Struktur zu bilden, welche nur die erlaubten beziehungsweise möglichen und sinnvollen Kombinationen beinhaltet

Hier ist eine Liste mit den verschiedenen Kombinationen:

- Filter/INPUT: Hier landen alle Pakete, die an einen lokalen Prozess gerichtet sind. Damit lassen sich Zugriffe auf lokale Prozesse perfekt regulieren, z.B.:
  -Zugriff auf einen lokal laufenden Server nur aus bestimmten Netzen
  -nur Pakete durchlassen, die zu einer bestehenden Verbindung gehören
- Filter/OUTPUT: Hier gehen alle Pakete durch, die von einem lokalen Prozess erzeugt wurden. Damit lassen sich lokale Prozesse nach aussen schützen, z.B.:
  -keine ausgehenden "verdächtigen" Verbindungen am Server -
  -keine "losen" Pakete nach draussen -- nur gültige Verbindungen
- Filter/FORWARD: Durch diese Chain gehen alle Pakete durch, die durch diese Maschine geroutet werden. Hiermit lassen sich also alle Rechner in jeweils dem Zielnetz des Routing schützen, z.B.:
  -kein UDP nach aussen, ausser DNS
  -keine öffnenden Verbindungen nach innen
  -Pakete, die zu keiner Verbindung gehören, werden gefiltert
- Nat/PREROUTING: Wenn Adress-Übersetzungen durchgeführt werden, müssen alle Pakete vor dem Routing hier durch. Hier lassen sich für zu routende Pakete verändern:
  -die Ziel-IP-Adresse
  -der Ziel-Port
- Nat/OUTPUT: Vom lokalen Rechner stammende Pakete gehen hier durch; gleiche Veränderungen wie Nat/PREROUTING.
- Nat/POSTROUTING: Hier gehen nochmals alle Pakete durch (auch lokal erzeugte Pakete) die geroutet worden sind. Hier werden Angaben über die Herkunft eines Paketes verändert, wie:
  -Quell-IP-Adresse
  -Masquerading (Sonderform von Quell-IP-Änderung)
- Mangle/PREROUTING & Mangle/OUTPUT:
  ähnlich den "nat" chains, nur mit dem Unterschied, dass hier spezielle Paket-Parameter geändert werden können, wie:
  - die TTL (Time to live)

(Vgl. Thüring, S.12f)

### 4.2.4 Statefull Firewalling

Der grösste Fortschritt ist sicher das "connection tracking". Netfilter überwacht Verbindungen und ordnet ihnen einen der vier Verbindungszustände NEW, ESTABLISHED, RELATED und INVALID zu. Die Erkennung von nicht zu einer Verbindung gehörenden Paketen und ihre Validierung durch den Kernel erhöhen die Sicherheit beträchtlich. Es muss also nicht mehr wie bei der Vorgängerversion alle Ports > 1024 aufgemacht werden, wenn nicht genau definiert ist, wo zum Beispiel die Antwortpakete von einem bestimmten Service hinein kommen. In der INPUT-Chain Regel wird einfach nur noch definiert, dass nur Pakete von aussen hereingelassen werden dürfen, die auch zu einer von innen aufgebauten Verbindung gehören.

# 5. Grenzen von Firewalls

Dieser Abschnitt soll erläutern, vor welchen Gefahren auch Firewalls nicht schützen können:

- Bei Angreifern aus dem internen Netz kann die Firewall nichts ausrichten. Wenn die Daten nicht über die Firewall geroutet werden, ist diese natürlich machtlos. Ein mögliches Szenario für diesen Fall wäre z.B. das ein Angreifer schon Zuggang zu einem Rechner im internen Netz hat. Jetzt kann er neue Angriffe auf andere Rechner im internen Netz starten, die alle nicht die Firewall passieren müssen, und deshalb erst mal unerkannt und unverhindert bleiben.
- Eine weitere Bedrohung sind Viren. Diese werden meist in ausführbaren Dateien oder eMails von Rechner zu Rechner übertragen und sind somit für die Firewall nicht automatisch zu erkennen. Aber auch hier gilt wieder einschränkend zu sagen, dass manche Virenangriffe (besonders von Würmern via http) zu verhindern sind, wenn der angreifende Virus oder Wurm eine eindeutige Signatur hat, also von der Firewall erkannt werden kann. Hierzu gehören Merkmale wie z.B. die URL, die aufgerufen wird.
- Eine weitere Angriffstaktik, gegen die Firewalls nichts ausrichten können, ist das so genannte Social Engineering . Dabei werden Angestellte vom Angreifer auf geschickte Weise ausgehorcht, in dem der Angreifer zum Beispiel anruft und vorgibt aus der Firma zu sein (bei großen Firmen sehr einfach) und nach Einstellungen, Passwörtern, Geburtstagen oder ähnlichen Daten fragt, die bei einem Einbruch hilfreich sein könnten. Da der Angreifer sich nachher eventuell bei der Firewall mit den

gewonnenen Daten authentifizieren kann, wird er nicht als Eindringling erkannt. (Vgl. Hardt/Tümmel S.18f)

# 6. Andere Sicherheitssysteme

Nachfolgend möchte ich kurz andere Sicherheitssysteme nennen. Diese sind z.T. in einer Firewall bereits implementiert.

- Intrusion Detection Systeme:
  Sie dienen der Erkennung von Angriffen. Dies geschieht z.B. durch Untersuchung von rohen Datenpaketen auf bekannte Angriffsmuster oder durch Untersuchung von Systemprotokollen.
- Scanner:
  Dies sind Erkennungstools für Sicherheitslücken. Der Scanner sucht dabei nach Schwachpunkten (aus einer Datenbank bekannt). Beispielsweise gibt ein Portscanner eine Liste mit allen offenen Ports zurück
- Protokollierung:
  Sie dient der Aufdeckung von Anomalien im Netzwerk und später zum Beweisen von betimmten Sachverhalten vor Gericht. Es ist darauf zu achten, dass der Cracker die Protokolldateien nicht modifizieren kann. Die Protokollierung ist mehr oder weniger in einer Firewall intergriert.

# 7.Trends

## 7.1 Firewalls basierend auf künstlicher Intelligenz

Die Grundlagenforschung im Bereich Computersicherheit beschäftigt sich seit einiger Zeit auch mit der Einbeziehung von Expertenssystemen in die Architektur von Firewall-Systemen. Dem Szenario von sich ständig ändernden Angriffsmethoden und neu hinzukommenden Sicherheitslücken wird versucht, mit Angriffserkennungsmethoden, die selbstlernende Strukturen wie neuronale Netzwerke oder Entscheidungsbäume

benutzen, zu begegnen. Erste Prototypen von intelligenten Firewalls wurden im Labormaßstab bereits getestet. (Vgl. Kyas, S.327)

## 7.2 Post-Firewall Ära

In welchem Ausmaß Firewall-Syseme auch in Zukunft als zentrale, für ein gesamtes Netzwerk zuständige Einheit realisiert werden müssen, läßt sich heute noch nicht abschätzen. Mit der zunehmenden Leistungsfähigkeit der Prozessoren rückt jedenfalls das Szenario von Comtputer-Systemen, die genügend Reserven an Rechenleistung besitzen, um sich selbst mit Hilfe von leistungsfähigen Authentifikations- und Verschlüsselungsmechanismen zu schützen, näher. Jedes einzelne System wird dann Daten ausschließlich über gesicherte Kommunikationskanäle mit identifizierten Kommunikationspartnern austauschen und so über ein privates Firewall-System verfügen, welches an Leistungsfähigkeit die heutigen System bei weitem übertrifft. (Vgl. Kyas, S.327f)

# 8. Fazit

Eine Firewall ist von unbedingter Notwendigkeit für die Erhöhung der Sicherheit. Eine gut administrierte Firewall ist in der Lage, die "normalen" (externen) Anwender von Spionage, Zerstörung von Dateien, u.a. wirkungsvoll abzuhalten. Aber dennoch ist ein Netzwerk auch mit der besten und teuersten Firewall (zumindest bis heute) noch nicht hundertprozentig sicher. Um die Sicherheit zu erhöhen sperrt man normalerweise Dienste und deinstalliert nicht benötigte Software, um die Anzahl der Sicherheitslücken zu verringern. Dabei muss man abwägen, in wieweit man diese Dienste beschränkt, um nicht den Nutzen und die Chancen, die die (weltweite) Vernetzung bringt, gänzlich wieder zunichte zu machen.

# Literaturverzeichnis

anonymous: Der neue Hacker's Guide, München 2001,
Markt + Technik-Verlag

Kyas Othmar: Sicherheit im Internet, 2. Auflage, Bonn 1998
MITP-Verlag

Pohlmann, Norbert: Firewall System, Bonn 2001, MITP-Verlag

Hardt Oliver und Tümmel Michael.: Seminararbeit "Firewalls",
Universität Hamburg

Thüring, Roman: Seminararbeit "Die SuSE Linux 7.3 Firewall"
Fachhochschule Aargau, 2002

Hegering H.G.: Praktikumunterlagen IT-Sicherheit, Universität
München, 2002

# Abbildungsverzeichnis

Schreiber, Alexander: Seminararbeit "Linux 2.4 Netfilter/iptables"
TU Chemnitz, 2000

**Universität Ulm**
**DaimlerChrysler**

# Seminar Internet-Dienste

Thema: Anonymisierung

Betreuer: Thilo Maier (DaimlerChrysler)

Johannes Bauer
Matr.-Nr.:431 270

27. Juli 2003

# Inhaltsverzeichnis

**1 Der gläserne Internetnutzer? 3**
1.1 Datenspeicherung auf Web-Servern . . . . . 4
1.2 Technische Weiterentwicklungen . . . . . 5

**2 Anonymes Surfen 8**
2.1 Einfache Proxyserver . . . . . 8
2.1.1 Funktionsweise . . . . . 8
2.1.2 Gefahren/Nachteile . . . . . 9
2.2 Web Mixe am Beispiel AN.ON der TU-Dresden . . . . . 11
2.2.1 Funktionsweise . . . . . 11
2.2.2 Mögliche Angriffspunkte . . . . . 14

**3 Zusammenfassung 17**

# 1 Der gläserne Internetnutzer?

Heutzutage besitzt der Großteil der PC-Nutzer einen Internetzugang oder ist zumindest an der Arbeitsstelle online. Zwei Hauptgründe für die allgemeine Akzeptanz des World Wide Webs dürften die einfache Bedienung und die ungeheure Menge an Informationen sein. Eine Internetverbindung aufzubauen gelingt selbst dem Computerlaien in wenigen Minuten, da die PCs von Discountern schon mit vorinstalliertem Browser und Modem geliefert werden. Fährt man den PC das erste Mal hoch, öffnet sich auch schon ein Fenster zur Erstellung eines Accounts bei einem Access-Provider.

Viele der Nutzer wissen jedoch nicht, dass sie im Internet mit der mitgelieferten Standardsoftware und den Browsereinstellung eine zum Teil breite Datenspur im Netz hinterlassen. Wie man anhand der Abbildung 1 erkennen kann, durchläuft ein Webseitenaufruf zahlreiche Instanzen [2]:

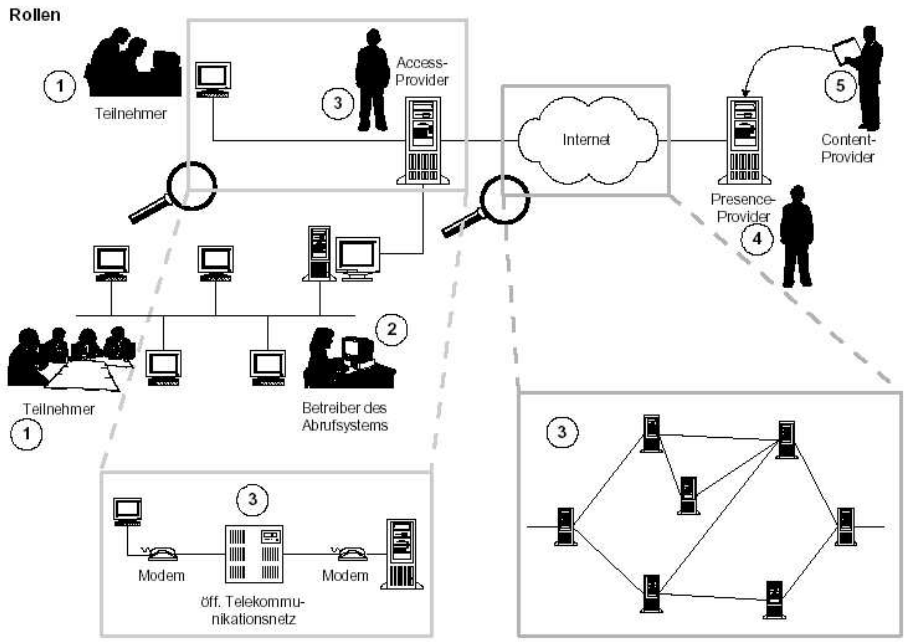



Abbildung 1: Datenfluss
Quelle: [2]

1. Der Teilnehmer
   besitzt einen lokalen Rechner (PC oder Workstation), mit dem er verschiedene Dienste des Internets nutzen kann. Dazu benötigt er mindestens den Zugangsvermittler und Dienstanbieter.

2. Der Betreiber des lokalen Rechners
   hat auf den Rechner des Teilnehmers ebenfalls teilweisen bis vollen

Zugriff. Er ist für die Funktion des lokalen Netzwerkes und die Anbindung an das Internet sowie Rechtevergabe zuständig. Beispiele sind der Systemadministrator eines Unternehmens oder Internetcafes.

3. Der Zugangsvermittler (Access-Provider)
   stellt die technische Infrastruktur für den Teilnehmer, bzw. Betreiber des lokalen Netzwerkes zur Verfügung um Dienste wie WWW, Email, usw. zu nutzen. Dazu gehören Telefonnetzbetreiber, Zugangsvermittler, Proxybetreiber, DNS-Server. Der Teilnehmer wählt sich über das Telefonnetz beim Zugangsvermittler ein und erhält von diesem eine IP-Adresse. Der DNS-Server dient zur Auflösung von textbezogenen Internetadressen in IP-Adressen.

4. Dienstanbieter (Presence-Provider, Service-Provider)
   Dienstanbieter verfügen über einen direkten Zugang zum Internet. Sie bieten ihren Kunden eigene Accounts an, die spezielle persönliche Leistungen ermöglichen. Dazu gehören z.B. Email-Verwaltung (GMX, MSN, ...), Internetzugänge (T-Online, AOL, 1und1, ...), Webhoster (Geocities, Freenet, ...) und viele andere Dienste. Häufige kombinieren die Dienstanbieter die verschiedenen Services und bieten ganze Pakete an.

5. Inhaltsanbieter (Content-Provider)
   ist jeder, der anderen zugängliche Informationen im Internet zur Verfügung stellt. Dies erfordert generell die Einbindung eines Dienstanbieters, welcher für den nötigen Speicherplatz und Internetzugang sorgt.

## 1.1 Datenspeicherung auf Web-Servern

Der Internetnutzer hinterlässt auf den angesurften Web-Servern Spuren, welche ihn im Normalfall eindeutig identifizieren und Rückschlüsse auf zuvor besuchte Seiten zulassen.

**IP-Adresse** Die IP-Adresse lässt eine eindeutige Unterscheidung aller direkt am Netz angeschlossenen Rechnern zu. Unterscheiden muss man zwischen statischen und dynamischen IP-Adressen.

Dynamische IP-Adressen werden vom Internet-Service-Provider (ISP) bei der Einwahl in das Internet zugewiesen. Bei jeder neuen Einwahl erhält der Nutzer eine andere Adresse aus dem Pool der Adressen des ISP. Auch können dynamische IP-Adressen während der Sitzung wechseln [4].

Dagegen sind statische IP-Adressen fest einem einzigen Nutzer zugewiesen. Diese sind v.a. bei Institutionen zu finden. Privatpersonen werden über ihren ISP meistens dynamische IP-Adressen zugeteilt, da insgesamt zu wenig IP-Adressen vorhanden sind. Ändern wird sich dies erst, wenn das neue Protokoll IPv6 eingeführt wird. Dort hat die IP-Adresse eine Länge von 128 Bit anstatt der derzeitigen von 32 Bit des IPv4 Protokolls.

Beobachtbar ist man folglich über dynamische IP-Adressen nur durch den ISP, da dieser als einziger eine Auflistung der in letzter Zeit genutzten Adressen besitzt. Üblicherweise speichert der ISP die IP-Adressen solange in einer Log-Datei, wie die Abrechnugsdaten des Kunden gespeichert bleiben. Dies sind im Regelfall einige Monate [2].

**Referrer** Der Referrer ist eine Variable, in der steht welche Internetseite der Surfer zuvor besucht hat. Dadurch können von Website-Betreibern gezielt Zugriffe, die von einer Konkurrenzseite erfolgen, blockiert werden. Gängiger dürfte allerdings die Auswertung sein, von welcher Seite der User kommt. Folgende Probleme können dadurch auftreten [2]:

- Die URL wird übertragen, dadurch werden auch interne bzw. geheime Adressen ersichtlich.
- Wird eine Seite über eine Suchmaschine aufgerufen, so wird somit auch die Suchanfrage übermittelt. Der Website-Betreiber kann also herausfinden, durch welche Schlüsselwörter und Interessen der User auf seine Seite aufmerksam geworden ist.

## 1.2 Technische Weiterentwicklungen

Nicht nur auf Webservern werden Daten gesammelt. Eine technische Weiterentwicklung sind Cookies, welche auf der Festplatte des Nutzers gespeichert sind und dadurch eine Wiedererkennung mit neuer IP-Adresse zulassen. Durch die Platzierung von Werbebannern und Web-Bugs, ist zudem die Generierung von Log-File-Einträgen auf nicht direkt besuchten Web-Servern möglich, da diese meist auf dem Server der Werbefirma gespeichert sind und nicht auf dem vom Content-Provider genutzten Server. Durch die Aktivierung von Java(Script) und ActiveX, können auf dem Rechner des Users Maschinencode und kleine Programme ausgeführt werden.

**Cookies** Cookies sind dafür verantwortlich, dass Sie auf einer zuvor besuchten Seite wiedererkannt werden. Dadurch entsteht auch der Aha-Effekt, wenn man z.B. persönlich mit dem eigenen Namen begrüßt wird.

Technisch betrachtet ist ein Cookie nur eine kleine Textdatei, die in einem bestimmten Verzeichnis des Rechners gespeichert ist. Ein Unterscheidungsmerkmal von Cookies ist die Lebensdauer. Diese kann mit dem Beenden des Browsers ablaufen (nicht persistent) oder durch eine im Cookie festgelegte Gültigkeitsdauer (persistent).

Der Inhalt eines Cookies sieht wie folgt aus [4]:

1. Domain: Domain, die das Cookie gesetzt hat und lesen darf.

2. Flag: Variable, die angibt, ob alle Computer der Domain Zugriff haben.
3. Path: Pfad der Domain, in der das Cookie gültig ist.
4. Secure: Variable, die angibt, ob nur über SSL auf das Cookie zugegriffen werden kann.
5. Expiration: Verfallsdatum des Cookies
6. Name: Name des Cookies
7. Value: Wert des Cookies (z.B. Identifikationsnummer)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| .spiegel.de | TRUE | / | FALSE | 1293839987 | RMID | c25fb01e38a00060 |
| .doubleclick.net | TRUE | / | FALSE | 1920499179 | id | dd3db276 |
| .zdnet.com | TRUE | / | FALSE | 1041310798 | cgversion | 4 |
| .zdnet.com | TRUE | / | FALSE | 1041310798 | browser | C25FB01E38A2A9E6 |
| .yahoo.com | TRUE | / | FALSE | 1271361599 | B | 03m4d1510tv9m&f=g |

Abbildung 2: cookies.txt Beispiel-Datei
Quelle:[4]

Genutzt werden Cookies, um Werbebanner gezielt einzublenden. Wenn man in das Cookie-Verzeichnis schaut, findet man mit großer Wahrscheinlichkeit ein Cookie der Internetwerbeagentur DoubleClick finden. Allein diese Firma hatte 1999 schon eine Datenbank mit 100 Millionen Dateien. Doubleclick ist auf vielen Seiten, v.a. Onlineshops, durch Werbebanner eingebunden und kann dadurch das Verhalten eines Users über eine eindeutige Identifikationsnummer, die im Cookie gespeichert ist, verfolgen. Angedacht hat DoubleClick auch eine Vernetzung ihrer Daten mit der der Direktmarketing-Firma Abacus Direct Corp. Somit könnte sogar ein Personenbezug hergestellt werden [7].

Einschränkungen im Cookie-Gebrauch können in den gängigen Browsern (InternetExplorer, Netscape, Mozilla) zwar ohne weiteres vorgenommen werden, jedoch stellt sich die Frage, ob dann noch die volle Funktionalität der Webseiten gewährleistet ist. Sperrt man die Cookies komplett, so werden sicher einige Seiten nicht mehr funktionieren; lässt man jedesmal per Pop-Up nachfragen, ob ein Cookie angenommen werden darf, wird man die Maustaste in nächster Zeit gehäuft benutzen. Eine sinnvolle Einstellung ist es, Cookies von Drittanbietern zu sperren, da diese nichts mit der im Browser dargestellten Seite zu tun haben. Bei den meisten Browsern besteht im Einstellungsmenü die Möglichkeit Cookies, die nicht von der Original-Webseite stammen, zu sperren. Dadurch wird verhindert, dass z.B. der Anbieter des auf der angewählten Ineternetseite angezeigten Werbebanners ein Cookie auf dem Rechner des Nutzers speichert.

**Werbebanner/Web-Bugs** Wie schon im Paragraph Cookies beschrieben, nutzen Firmen, wie z.B. DoubleClick Werbebanner, um das Nutzungsverhalten einer Person über mehrere Seiten nachzuverfolgen. Daraus kann dann ein spezifisches Profil erstellt werden, welches für die gezielte Einblendung von Werbebannern genutzt wird.

Diese Profilerstellung kann jedoch auch viel unauffälliger erfolgen. Dazu bedienen sich solche Firmen den sogenannten Web-Bugs. Web-Bugs sind Grafikreferenzen auf 1x1 Pixel große Bilder, die in e-Mails mit HTML oder auf Webseiten mit einem einfachen HTML-Kommando referenziert werden [6]:

*<img src="http://www.dritter.de/anbieter.gif" width=1 height=1 border=0>*

Besonders gut eignen sich hierfür GIF-Dateien, da diese transparent gestaltet werden können und somit für das Auge unsichtbar sind. Da der Web-Bug auf dem Server des Drittanbieters gespeichert ist, bekommt dieser bei jedem Zugriff auf die Webseite des Content-Providers ebenfalls einen Eintrag in seine Log-Datei.

**Java(Script)/ActiveX** Die nachfolgende Tabelle soll die Schwachpunkte der Programmiersprachen JavaScript, Java und ActiveX in Zusammenhang mit dem Gebrauch im Internet darstellen [8]:

| | **JavaScript** | **Java** | **ActiveX** |
|---|---|---|---|
| **Entwickler** | Netscape | Sun Microsystems | Microsoft |
| **Zugriff auf Systemressourcen** | nein | eingeschränkt | vollständig |
| **Risiken** | Implementierungsfehler | Implementierungsfehler | Zugriff auf alle Sytemressourcen |
| **Ausführung** | durch Browser | Java Virtual Machine | Maschinencode /JVM |

Durch den teilweise möglichen Zugriff auf Systemressourcen können Programme ausgeführt werden, die die IP-Adresse oder andere vertrauliche Daten übermitteln. ActiveX eröffnet hier dem Angreifer die besten Möglichkeiten.

# 2 Anonymes Surfen

Eine der wichtigsten Voraussetzungen um unerkannt und unbeobachtbar im Internet surfen zu können besteht darin, nicht durch die IP-Adresse eindeutig indentifizierbar zu sein. Der Folgende Teil befasst sich mit dem grundsätzlichen Verfahren der Proxyserver, im Speziellen wird der Anonymisierungsdienst der TU-Dresden betrachtet.

## 2.1 Einfache Proxyserver

Einfache Proxyserver zeichnen sich dadurch aus, dass sie zwar die IP-Adresse des Anwenders unterdrücken, aber ihre eigentliche Intention nicht alleine auf Sicherheitsgründen aufgebaut ist, sondern auch Überlegungen der Geschwindigkeitsperformance oder des Bandbreitenverbrauchs berücksichtigt. Es ist bei diesen Servern nicht nötig zusätzliche Anonymisierungsprogramme auf dem PC des Anwenders zu installieren. Die Adresse des Proxyservers wird in den Browsereinstellungen nur in das entsprechende Feld eingetragen. Bei formularbasierten Proxyservern ist nur ein einfacher Browser nötig (Beispiel: www.anonymizer.com).

### 2.1.1 Funktionsweise

Proxy heißt auf deutsch "Bevollmächtigter" oder "Stellvertreter" [10]. Dies beschreibt schon einen Teil seiner Funktionsweise. Der Proxy ist dazu da, Nutzeranfragen nicht direkt an den Dienstanbieter zu stellen, sondern diese zuerst an den Proxy-Server zu leiten. Dieser leitet die Anfrage stellvertretend an den Dienstanbieter weiter, holt die benötigten Informationen und stellt diese dann dem Nutzer zur Verfügung. Der Proxy erscheint gegenüber dem Nutzer als Gateway, während dieser gegenüber dem eigentlichen Webserver als Client auftritt.

Die Proxyserver unterscheiden sich hauptsächlich in der Verwendung als Cache-Proxies oder Proxies ohne Cache. Der Cache ist ein Speicher, welcher verwendet wird um abgerufene Inhalte zwischenzuspeichern und diese bei zeitnahem Abruf durch einen anderen Client direkt an diesen weiterzugeben, um die Requestzeit und Datenübertragung zum Webserver einzusparen.

Die wesentlichen Gründe für den Betrieb von Proxyservern sind folgende [10]:

- Zugriffsteuerung
  In Firmen- und Uninetzen werden Proxies verwendet um Zugriff auf bestimmte Teile des Internets zu untersagen, z.B. Seiten mit pornografischen oder rechtswiedrigen Inhalten.
- Sicherheit
  Der Client tritt durch die Umleitung über den Proxy nicht selber mit

seiner IP-Adresse in Erscheinung. Manche Proxyserver filtern auch potenziell gefährliche Daten heraus (Cookies, Skriptcode, ...). Problem ist hier dann jedoch die unter Umständen ungenügende Darstellung von seriösen Internetseiten, wie z.B. Internetbanking.

- Protokollierung
  Wie auch bei normalen Webservern wird der Zugriff auf den Proxy mitprotokolliert und kann natürlich auch mit entsprechenden Programmen ausgewertet werden.

- Beschleunigung und Minimierung der Datenübertragung
  Durch den Einsatz von Cache-Proxies kann die Übertragung von Daten aus dem Internet durch die kürzere Distanz zum Proxy beschleunigt werden und durch die lokale Speicherung dadurch noch Bandbreite eingespart werden. Das Prinzip dieses Aspektes entspricht etwa dem des Browser-Caches.

Proxyserver die vor allem aus Gründen der "Sicherheit" bestehen nennt man ANON-Proxies (ANON vom englischen Wort anonymity). Sie dienen dem Zweck der Unterdrückung der IP-Adresse des Clients. Ob ein Proxy die IP-Adresse nicht an den Webserver weiterleitet kann man leicht überprüfen, indem man eine Seite mit IP-Check aufruft (z. B. http://checkip.dyndns.org/). Google liefert hier etliche Anbieter, die ebenfalls noch überprüfen, welche zusätzlichen Informationen (Betriebssystem, Sprache, Browser, installierte Plugins, ...) der Browser überträgt.

### 2.1.2 Gefahren/Nachteile

(ANON-)Proxyserver fallen durch ihre unkomplizierte Benutzbarkeit auf, da ein einfacher Eintrag im Browser schon ausreicht um "sicher" und "unbeobachtbar" zu surfen. Das größte Sicherheitsrisiko besteht darin, dass die Anonymisierung erst auf dem Server des Anonymitätsdienstes beginnt. Für den Betreiber des Dienstes ist es kein Problem, die Aktivitäten jedes Nutzers in den Logfiles mitzuprotokollieren und gegebenenfalls für wirtschaftliche Zwecke auszuwerten. Für Datensammler wie DoubleClick oder ähnliche Firmen wäre es leicht, einen ANON-Proxy zu installieren oder aufzukaufen, um so an unverfälschte IP-Adressen zu gelangen. Ein zweiter Vorteil für solche Dienste wäre, dass sie das komplette Surfverhalten eines Nutzers überwachen können, da der gesamte Verkehr über den Proxybetreiber läuft. Die Übliche Auswertung von Web-Bugs, Werbebannern usw. lässt keine vollständige Profilerstellung zu. Es kommt also in erster Linie darauf an, dem Serverbetreiber zu vertrauen.

Ein weiterer wichtiger Punkt sind vielfach verwendete Programmiersprachen auf Webseiten und die Speicherung von Cookies. Dort muss beim Serverbetreiber entschieden werden, welche Codefragmente bzw. Cookies her-

ausgefiltert werden und welche zu Gunsten des Erscheinungsbildes an den Teilnehmer weitergegeben werden. Da sich diese Sprachen und auch der HTML-Code ständig weiterentwickeln, können auch vermeintlich sichere Proxies unter Umständen durch nicht vollständig geparsten Code die IP-Adresse des Teilnehmers nicht verstecken. Als Beispiel dienen folgende HTML- und StyleSheet-Dateien:

liste.html:

```
<html>
<head>
<link rel="stylesheet" href="style.css">
</head>
<body>
Liste:
<ul>
   <li>Punkt 1
   <li>Punkt 2
</ul>
</body>
</html>
```

style.css:

```
ul {list-style-image:url(bild.gif);}
```

Die Anweisung in *style.css* bewirkt, dass die in der *liste.html* mit <ul>-Tag generierten Listen nicht mit Punkten, sondern mit dem Bild dash.gif dargestellt werden. Anfang 2000 konnte der neueste Internet-Explorer dieses neue Feature zwar darstellen, doch einige formularbasierte Anonymitätsdienste parsten diesen Teil der *.css Datei nicht. Dadurch wurde das Bild *bild.gif* vom Browser des Nutzers geladen und dadurch die IP-Adresse übermittelt [11].

Eine dritte Schwäche von ANON-Proxies ist die mögliche Beobachtung durch einen sogenannten Big Brother. Als Big Brother bezeichnet man einen starken externen Angreifer, der große Teile des Netzwerkes überwacht. Dabei ist es nicht zwingend erforderlich das gesamte WWW zu kontrollieren, es genügt wenn der Angreifer den für ihn interessanten lokalen Bereich unter Kontrolle hat. Dies könnte z. B. ein totalitärer Staat sein, der das gesamte Kommunikationsnetz des Landes überwacht. Obwohl er nicht der Proxybetreiber ist, kann er durch die Kontrolle aller aus- und eingehenden Datenströme des Proxybetreibers Aufschluss über einen Nutzer bekommen. Er kann bei verschlüsselter Kommunikation zwar nicht direkt die übertragenen Inhalte auslesen, jedoch ist für den Big Brother ersichtlich welcher Webserver von welchem Benutzer aufgerufen wurde. Dies wird ihm durch die zeitnahe Übertragung der Daten ermöglicht. Da er genau weiß, wann von einem

betimmten Nutzer der Request am Proxyserver eingeht, muss der nächste ausgehende Request eben dieser sein. Da dieser wiederum über das staatliche Kommunikationssystem an den Webserver weitergeleitet wird, erfährt der Big Brother nun, welche Seite der Internetnutzer abgerufen hat.

## 2.2 Web Mixe am Beispiel AN.ON der TU-Dresden

Anonymity online (AN.ON) ist ein Projekt der technischen Universität Dresden, welches von der Deutschen Forschungsgemeinschaft und vom Bundesministerium für Wirtschaft und Technologie (BmWi) unterstützt wird. Das Ziel des Projektes ist die Ermöglichung des anonymen Zugriffs auf das Internet für jeden User.

### 2.2.1 Funktionsweise

Im Gegensatz zu den einfachen Proxyservern wird bei AN.ON die Möglichkeit eines Big Brother Angriffes als realistisch betrachtet. Zwar befindet sich das Projekt noch im Entwicklungsstadium, doch ist der technische Aufbau so ausgelegt, dass im Endstadium auch starke "allesüberwachende" Angreifer keine, bzw. sehr geringe Chancen der Kommunikationsüberwachung bestimmter oder aller Nutzer haben.

Die Idee der Web-Mixe stammt von David Chaum, einem amerikanischen Kryptographen, und wurde bereits in den 80er Jahren von ihm zum anonymen Senden und Empfangen von E-Mails entwickelt [15]. Das grundlegende Prinzip der Web-Mixe besteht aus der Hintereinanderschaltung mehrerer Anon-Proxies. Verschlüsselte Daten von vielen Nutzern werden in den Mixen durcheinandergemischt, neu verschlüsselt, an den nächsten Mix einer Kette weitergegeben und beim letzten Mix entschlüsselt [13]. Dieser tätigt die Anfrage beim Webserver und schickt die angefragten Daten in gleicher Weise über die Mix-Kette zurück. Dadurch verliert selbst ein Big Brother die Zuordnungsmöglichkeit der Datenpakete zu den Nutzern, solange ein Mix nicht durch ihn kontrolliert wird.

AN.ON besteht aus mehreren Komponenten welche in Abbildung 3 dargestellt sind.

**JAP** Der für den Internetnutzer wichtigste Teil ist der JAP (Java Anon Proxy), ein Java Programm, welches unter http://anon.inf.tu-dresden.de/ momentan noch kostenlos zu beziehen ist. Durch die Verwendung von Java als Programmiersprache ist ein plattformübergreifender Einsatz leicht möglich [13]. Diese Software muss auf dem Rechner des Nutzers installiert werden. Der JAP übernimmt dort die Funktion eines Proxies, der die Verbindung zum ersten Mix herstellt. Damit ist es theoretisch möglich alle proxyfähigen Dienste zu nutzen. In der Praxis werden jedoch nur Anwendungen unterstützt welche über den Browser laufen, um eine zu hohe Datenflut durch

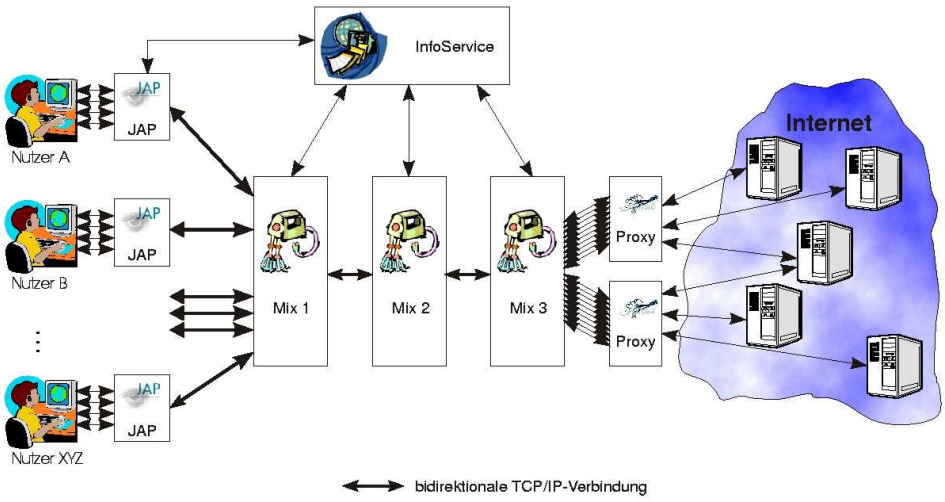



Abbildung 3: AN.ON
Quelle: [9]

Tauschbörsen wie z.B. Kazaa, eDonkey und eMule zu verhindern. Die zweite Funktion des JAPs ist die mehrfache Verschlüsselung der zu sendenden Daten.

Um die Aktivität des Nutzers geheimzuhalten ist es möglich sogenannten "Dummy-Traffic" zu aktivieren. Dieser ebenfalls verschlüsselte "sinnlose" Verkehr kann nicht von echten Anfragen unterschieden werden. Senden alle Nutzer des Anonymitätsdienstes diese Datenpakete durch die Kaskade, ist für einen Überwacher nicht ersichtlich, welcher Internetnutzer gerade akitv ist und welcher nicht.

**Mixe** Die durch den JAP mehrfach verschlüsselten Daten, werden anschließend durch eine wählbare Mix-Anreihung gesendet. Der Unterschied zu dem Konzept von Chaum besteht hier in der Anordnung der Mixe. Chaum stellt in seinem Konzept eine Anzahl frei wählbarer Mixe zur Verfügung. Der Nutzer kann selbst die Mixe bestimmen, die seine Nachrichten durchlaufen sollen und auch deren Reihenfolge. Ein Mix ist ein Proxyserver mit den später beschrieben Funktionen.

Das AN.ON-Projekt basiert auf Mix-Kaskaden. Eine Mix-Kaskade ist eine Hintereinanderreihung von Mixen. Mix-Kaskaden erlauben dem Nutzer lediglich die Wahl einer solchen Kaskade. Die Server der Kaskade werden durch den Betreiber (hier: TU-Dresden) vorgegeben. Der User wählt daraus die, die ihm am vertrauenswürdigsten erscheint. Der Vorteil von Mix-Kaskaden gegenüber frei wählbaren Mix-Netzwerken ist, dass dadurch die

Anonymitätsgruppe (Gruppe, die eine bestimmte Mixanreihung wählt) der Anwender, die die gleiche Abfolge von Mixen wählen, größer wird. Dadurch wird es dem potentiellen Angreifer erschwert sogenannte Schnittmengenangriffe zu starten [14]. Als Schnittmengenangriffe bezeichnet man die langfristige Beobachtung der Online- und Offline-Phasen von Internetnutzern und dessen Gewohnheiten, um seine Aktivitäten zu isolieren. Je größer die Anonymitätsgruppe ist, desto schwerer ist die Verkettung einzelner Aktivitäten mit einer bestimmten Person. Aus Performancegründen ist jeder Mix nur in einer Kaskade vorhanden [12].

Um die Anonymität der Nutzer zu gewährleisten verpflichtet sich jeder Mix-Betreiber, keinerlei Daten mit anderen Mix-Betreibern auszutauschen und keine Log-Dateien anzulegen. Dieses Versprechen nützt jedoch nichts, falls der Angreifer der Betreiber eines Mixes ist.

Um trotzdem Anonymität zu gewährleisten, werden die Daten des Nutzers schon auf dessen Rechner verschlüsselt und zwar genau so viele Male, wie es Mixe in der gewählten Kaskade gibt. Die Verschlüsselung beginnt beim Nutzer für den letzten Mix der Kaskade, dann nochmals für den Vorletzten bis hin zum Ersten. Diese Nachricht wird dann an den ersten Mix gesendet, der sie entschlüsselt und an den zweiten Mix in der Kette weitersendet, welcher wieder entschlüsselt, bis sie beim letzten Mix vollständig entschlüsselt wird. Jeder Mix kann nur die für ihn vorgesehene Verschlüsselungsstufe entschlüsseln. Die Entschlüsselung an den unterschiedlichen Mixen hat zur Folge, dass die Nachricht nicht verfolgt werden kann, da sich bei jedem Entschlüsseln die Struktur verändert. Dadurch kennt jeder Mix nur seinen Vorgänger und den Nachfolger, nie aber gleichzeitig Sender und Endadressat (Webserver) der Datenpakete, sofern die Kaskade aus wenigstens 2 Mixen besteht.

Um eine Analyse der Datenströme durch einen Big Brother zu verhindern, hat jede Nachricht die gleiche Länge. Außerdem verarbeiten die Mixe die Daten schubweise, d.h. sie sammeln erst eine gewisse Anzahl an eingegangenen Datenpaketen, mischen diese durcheinander[1], und schicken sie dann zusammen an den nächsten Mix weiter [15]. Diese Vorgehensweise hat zur Folge, dass es genügt, wenn in der Mix-Kaskade mit mindestens zwei Mixen[2], nur ein einziger Mix nicht mit den anderen kooperiert, um unbeobachtbares Surfen zu ermöglichen. Werden Mix-Kaskaden aus einer heterogenen Gruppe vertrauenswürdiger Betreiber gebildet (z.B. Behörden, Kirchen, soziale Institutionen, Universitäten, ...) ist die Wahrscheinlichkeit ziemlich hoch, dass wenigstens einer korrekt arbeitet.

**Infoservice** Der Infoservice versorgt den JAP mit allen nötigen Informationen, die für das Durchlaufen der Mixe notwendig sind [13]:

[1] daher stammt der Name Mix

[2] in der Praxis sind es zur Zeit drei

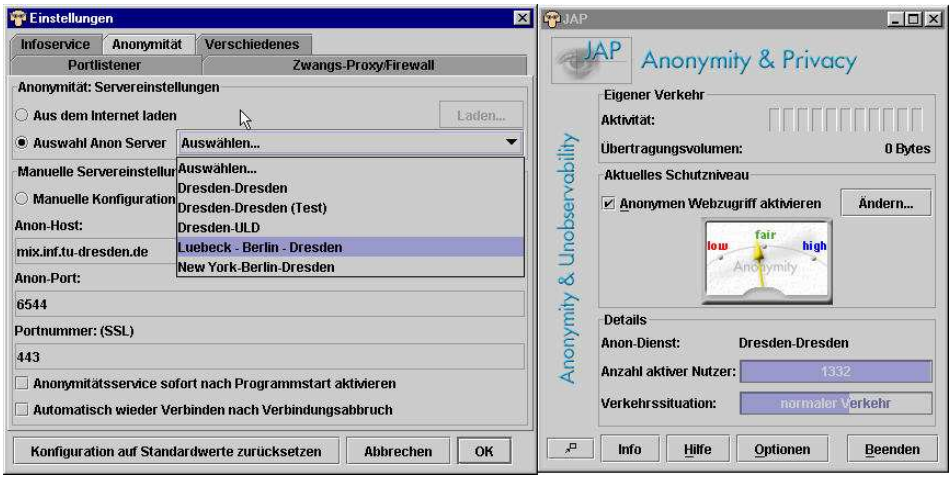



Abbildung 4: Infoservice

- Adressen und öffentliche Schlüssel der Mixe
- Menge der aktiven Nutzer und des Datenverkehrs
  Aus diesen beiden Zahlen wird ein Anonymitätsgrad des Nutzers berechnet. Generell gilt: je mehr verschiedene Nutzer und mit je höherer Datenübertragungsrate AN.ON gleichzeitig nutzen, desto sicherer ist das System.
- Verfügbarkeit der einzelnen Mix-Kaskaden

**Cache-Proxy** Der Cache-Proxy sendet die Requests an die Webserver, falls nicht im Cache schon vorhanden, und leitet die Antworten über die Mix-Kaskade in umgekehrter Reihenfolge zurück an den Internetnutzer. Dies geschieht wieder in gleich großen Datenpaketen [13].

### 2.2.2 Mögliche Angriffspunkte

Wie man anhand des Paragrafen "Mixe" sehen kann, hat alleinige Beobachtung der ein- und ausgehenden Datenpakete an allen Mixen keine Beinträchtigung der Anonymität zur Folge, solange wenigstens ein Mix die Zuordnung von ein und ausgehenden Paketen geheim hält.

Gefährlicher sind aktive Angreifer, die selbst Daten durch die Mixketten senden und dadurch versuchen einzelne Benutzer zu enttarnen.

**Replay-Angriff** Hier zeichnet der Beobachter eine Nachricht eines Benutzers auf und versucht durch erneutes Absenden der Nachricht nochmals die Mix-Kaskade zu durchlaufen. Da diese Nachricht zweimal durch den Mix gesendet wurde, ist es möglich durch Vergleichen der ein- und ausgehenden

Nachrichten eines Mixes die doppelt gesendete Nachricht herauszufiltern und dadurch den Mix zu überbrücken.

Deshalb erhält jede Nachricht im Mix einen eindeutigen Fingerabdruck, der in einer Datenbank gespeichert wird. Erhält der Mix eine Nachricht, prüft er zuerst, ob diese schon einmal bearbeitet wurde. Ist dies der Fall, wird die Nachricht gelöscht, ansonsten wird sie normal weiterbearbeitet. Aus Performancegründen enthält jede Nachricht eine begrenzte Gültigkeitsdauer. Wird sie nicht vor dem Ablaufen des Zeitstempels bearbeitet, wird sie ebenfalls gelöscht. Somit verhindert der Mix-Betreiber, dass die Datenbank mit den Fingerabdrücken zu groß wird [12, 15].

**"$n$-1"-Angriff** Der "$n$-1"-Angriff ist der stärkste anzunehmende Angriff. $n$ ist die Zahl der Nachrichten eines Schubes. Ein Schub ist die Ansammlung einer bestimmten Anzahl von Datenpaketen mit anschließender gleichzeitiger Weitersendung. Der Angreifer generiert innerhalb dieses Schubes $n$-1 Nachrichten, um die übrigbleibende Nachricht zu isolieren und damit verfolgen zu können. Dies kann natürlich nur durch sehr starke Angreifer, welche einen hohen Aufwand betreiben geschehen. Der Angreifer muss nicht nur genügend Nachrichten generieren damit der Schub ausgeführt wird, er muss auch in der Lage sein, sicherstellen zu können, dass kein anderer Nachrichten in diesem Moment sendet. Dies kann nur geschehen, wenn er die restlichen Nutzer kontrolliert, bzw. selbst soviele Nutzer simuliert, dass ein vermeintlich sicheres Anonymitätslevel vorgetäuscht wird.

Zur Abwehr dieses Angriffes gibt es einen Ansatz, welcher momentan noch nicht in das AN.ON-System der TU-Dresden implementiert ist. Jeder User muss sich zur eindeutigen digitalen Identifizierung bei einer oder mehreren Zertifizierungsstellen registrieren. Wichtig dabei ist, dass die Zertifizierungsstelle(n) vertrauenswürdig ist (sind) und jeder Person nur ein Zertifikat ausstellen. Mit diesem Zertifikat meldet sich der Nutzer am Mix an. Für jeden Schub in dem er eine Nachricht senden möchte erhält der Nutzer vom Mix ein Ticket, welches mit der zu sendenden Nachricht später versandt wird. Dadurch wird nachgewiesen, dass die Nachricht von einem autentifizierten Teilnehmer stammt [13].

Das Ticket muss folgende Eigenschaften aufweisen [16]:

- Nur der Mix darf gültige Tickets generieren.
- Das Ticket darf nur dem Nutzer und dem Mix bekannt sein.
- Das Ticket muss schubbezogen sein, d.h. es darf nur für einen bestimmten Schub gültig sein, um die Ansammlung von Tickets über mehrere Schübe zu verhindern.
- Der Mix darf keine Verbindung zwischen Ticket und Teilnehmer herstellen können.

- Die Ticketanzahl, die ein Nutzer während eines Schubes erhält muss begrenzt sein.

**Aktive Inhalte** Wie auch bei normalen (ANON-)Proxies ist es bei Mix-Kaskaden erforderlich aktive Inhalte wie Java(script), ActiveX und auch Cookies abzuschalten damit die IP-Adresse und das Nutzerverhalten geheim bleiben. Die Herausfilterung von aktiven Inhalten im Proxy stellt immer ein Risiko dar. Zudem überträgt die Mix-Kaskade und der Cache-Proxy sämtliche Inhalte. Da der Cache-Proxie des Projektes AN.ON keine Inhalte herausfiltert schlägt das JAP-Modul bei der Installation vor, die richtigen Einstellungen im Browser zu tätigen.

# 3 Zusammenfassung

Die meisten Nutzer hinterlassen im Internet (un)wissentlich eine Datenspur. Angefangen von Referrer über IP-Adresse und Cookies, bis hin zu ausgewerteten Klicks auf Werbebanner. Dadurch lässt sich mit ausreichendem Willen eines Beobachters ein möglicherweise ziemlich genaues Nutzungsprofil einer Person aufzeichnen. Vielerorts ist diese Vorgehensweise nicht unnützlich, (z.B. Amazon, Ebay) um über geeignete Produkte informiert zu werden. Jedoch sollte die letztendliche Entscheidung, wieviel ein Nutzer von sich preisgeben will, in seinem Ermessensspielraum liegen.

Dass das Verlangen nach Anonymität im Internet immer größer wird, ist an der mittlerweile großen Zahl an Anonymitätssdiensten abzuleiten. Sogar kommerzielle Dienste wie Anonymizer.com (http://www.anonymizer.com/) scheinen genügend Interessenten zu finden, die bereit sind für Ihre Anonymität zu bezahlen. Die Gefahr besteht, dass in der Zukunft auch Marketingagenturen eigene vermeintliche Anonymitätsdienste betreiben werden, um Internetnutzer noch genauer in Profile zu zwängen.

Möglichkeiten seine Identität zu verschleiern gibt es genügend. Die Frage ist nur, ob sich der Aufwand lohnt, bzw. ob man mit der eingeschränkten Funktionalität des Internets, dann noch einverstanden ist.

# Abbildungsverzeichnis


1 Datenfluss . . . . . 3
2 cookies.txt Beispiel-Datei . . . . . 6
3 AN.ON . . . . . 12
4 Infoservice . . . . . 14

## Literatur


[1] Hannes Federrath: Zur Kontrollierbarkeit des Internet, http://page.inf.fu-berlin.de/~feder/publ/1999/Fede2_99ZUM.pdf, 1998

[2] Marit Köhntopp, Kristian Köhntopp: Datenspuren im Internet, http://123.koehntopp.de/kris/artikel/datenspuren/CR_Datenspuren_im_Internet.pdf, 2000

[3] Hannes Federrath, Oliver Berthold, Marit Köhntopp, Stefan Köpsell: Tarnkappe für das Internet. Verfahren zur anonymen und unbeobachtbaren Kommunikation., c't 16/2000

[4] Bundesamt für Sicherheit in der Informationstechnik: Das Ende der Anonymität? Datenspuren in modernen Netzen., http://www.bsi.de/literat/anonym/index.htm

[5] Bundesamt für Sicherheit in der Informationstechnik: Ins Internet mit Sicherheit, http://www.bsi-fuer-buerger.de/druck/kap_02.pdf, Kapitel 2

[6] Michael Miller: Absolute PC Security and Privacy, Sybex Inc., Alameda, 2002, Chapter 24

[7] Simon Perkins: Internet Cookies: Security Implications, http://www.cs.uct.ac.za/courses/CS400W/NIS/papers00/sperkins/cookies.ps, Mai 2000

[8] Simson Garfinkel, Eugene H. Spafford: Web Security & Commerce, O'Reilly, 1997

[9] Hannes Federrath, Stefan Köpsell, Heinrich Langos: Anonyme und unbeobachtbare Kommunikation im Internet, http://page.inf.fu-berlin.de/~feder/publ/2002/FeKL2002GIJT.pdf, 2002

[10] Dirk Dithardt: Squid Proxyserver Administrationshandbuch -Entwurf-, http://www.squid-handbuch.de/, 2003

[11] Hannes Federrath: Unsicherheit formular-basierter Anonymitätsproxies, http://www.inf.tu-dresden.de/~hf2/anon/aproxies/, 2001

[12] Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein: Sicherheit im Internet durch Anonymität, http://www.datenschutzzentrum.de/selbstdatenschutz/anonym/, 2002

[13] Oliver Berthold, Hannes Federrath, Stefan Köpsell: WebMIXes: A System for anonymous and unobservable Internet access, http://www.inf.fu-berlin.de/~feder/publ/2001/BeFK2001BerkeleyLNCS2009.pdf, 2001

[14] George Danezis: Mix-networks with Restricted Routes, http://www.cl.cam.ac.uk/~gd216/ExpMix.pdf

[15] Technische Universität Dresden: Technischer Hintergrund von JAP, http://anon.inf.tu-dresden.de/JAPTechBgPaper.pdf

[16] Oliver Berthold, Hannes Federrath, Stefan Köpsell: Praktischer Schutz vor »Flooding-Angriffen« bei Chaumschen Mixen, http://www.inf.fu-berlin.de/~feder/publ/2001/BeFK2001KoSi.pdf, 2001

Seminarvortrag zum Thema

# Grid-Computing

Norbert Heidenbluth (Stand: 26.07.2003)

Seminar Internet-Dienste - Sommersemester 2003
Professor Dr. Franz Schweiggert - Universität Ulm

# Inhaltsverzeichnis

**1 Einführung** **4**

**2 Anwendungstypen des Grids** **7**

2.1 Begriffsdefinitionen . . . . . 9

2.2 Rechen-Grids . . . . . 9

2.2.1 Ausnutzen nicht verwendeter Ressourcen . . . . . 10

2.2.2 Parallele CPU-Kapazitäten . . . . . 11

2.2.3 Geeignete Aufgaben für verteiltes Supercomputing . . . 11

2.3 Data-Grids . . . . . 12

2.3.1 Temporärer vs. permanenter Speicherplatz . . . . . 12

2.3.2 Data Striping . . . . . 13

2.4 Kollaborative Computerarbeit . . . . . 14

2.4.1 Virtuelle Ressourcen und virtuelle Organisationen . . 14

2.4.2 Reservation, Scheduling und Scavenging . . . . . 15

2.4.3 Ausgleich von Resourcen . . . . . 17

2.4.4 Betriebssicherheit und Zuverlässigkeit . . . . . 18

**3 Aufbau und Anwendung eines Grids** **19**

3.1 Der stufenweise Aufbau eines Grids . . . . . 19

3.2 Grid Software Komponenten . . . . . 20

3.2.1 Management Komponenten . . . . . 20

3.2.2 Donor Software Komponenten . . . . . 21

3.2.3 Submission Software Komponenten . . . . . 21
3.2.4 Scheduling Software Komponenten . . . . . 21
3.2.5 Communication Software Komponenten . . . . . 21
3.3 Grid-Architektur . . . . . 22
3.3.1 Fabric Layer . . . . . 22
3.3.2 Connectivity Layer . . . . . 23
3.3.3 Resource Layer . . . . . 24
3.3.4 Collective Layer . . . . . 24
3.3.5 Applications Layer . . . . . 24
3.4 Das Globus Toolkit . . . . . 25
3.5 Die Verwendung eines Grids . . . . . 25
3.5.1 Schritt 1: Installation der Grid-Software . . . . . 25
3.5.2 Schritt 2: Anmeldung am Grid . . . . . 25
3.5.3 Schritt 3: Anfragen und Übermittlung von Jobs an das Grid . . . . . 26
3.5.4 Schritt 4: Statusüberwachung und Wiederherstellung . 26
3.5.5 Schritt 5: Ergebnisauswertung . . . . . 27
3.6 Sicherheitsaspekte . . . . . 27

**4 Beispiele für Grids** **29**
4.1 SETI@home . . . . . 29
4.1.1 Ursprung des Projekts . . . . . 29
4.1.2 Übermittlung der Daten . . . . . 30
4.1.3 Sicherheit und Zuverlässigkeit . . . . . 31
4.1.4 Bisheriger Erfolg . . . . . 32
4.1.5 Zu guter Letzt... . . . . 33
4.2 IBM ZetaGRID . . . . . 33
4.3 CERN-Data-Grid (LHC) . . . . . 35
4.4 Weitere Projekte . . . . . 35

**5 Abschließende Bemerkungen** **38**

5.1 Grid-Computing und seine Beziehungen zu anderen Technologien . . . . . 38

5.1.1 World Wide Web . . . . . 38

5.1.2 Application und Storage Service Providers . . . . . 38

5.1.3 Peer-to-Peer- und Internet-Computing . . . . . 39

5.2 Aussagen zum Grid-Computing und deren Bewertung . . . . . 39

5.3 Probleme des Grid-Computings . . . . . 40

# Kapitel 1

# Einführung

Stellen wir uns vor, unser Nachbar besitzt einen teuren und vor allem schnellen Sportwagen, um den ihn alle beneiden. Aber was macht er mit dem Fahrzeug? Er fährt damit morgens für drei Minuten zum Bäcker um die Ecke, kauft die Frühstückssemmeln und stellt den Wagen danach für den Rest des Tages wieder in seine Garage. Sicherlich werden die meisten Menschen denken, daß es doch "eine Schande" sei, die zur Verfügung stehende Leistung des Sportwagens so ungenutzt zu lassen.

Übertragen wir jetzt dieses Szenario von der Welt der Automobile auf die Welt der Computer, so stellen wir fest, daß fast jeder Computerbesitzer ebenfalls immense Kapazitäten über weite Teile des Tages ungenutzt auf seinem Schreibtisch stehen hat. Und an dieser Stelle setzt der Gedanke des Grid-Computing ein. Eine der grundlegendsten Anwendungen des Grid-Computings möchte die brachliegende Leistung von PCs, Workstations oder gar Servern nutzen, um sie für große Rechenprojekte zu verwenden, die auf einzelnen Computern gar nicht oder zumindest nicht in einer überschaubaren Zeit zu Ergebnissen führen würden. Und so zeigt sich Dr. Franz-Josef Pfreundt, Forscher am Fraunhofer Institut für Techno- und Wirtschaftsmathematik (ITWM) in Kaiserslautern, davon überzeugt, daß "Grid-Computing der Trend der Computerbranche" sei. [1].

Was aber verbirgt sich hinter dem Begriff "Grid"? Versuchen wir es zunächst mit einer technischen Definition:

> "Grid is a type of parallel and distributed system that enables the sharing, selection and aggregation of geographically distributed "autonomous" resources dynamically at runtime depending on their availability, capability, performance, cost, and users' quality-of-service requirements."[2]

Im weiteren Verlauf dieser Seminararbeit werden wir die einzelnen Begriffe, die obiger Definition zugrunde liegen, noch näher erläutern. Bis dies jedoch geschehen ist, nähern wir uns dem Begriff Grid-Computing mit einer weiteren Erklärung:

> "Einer der wichtigsten Erfolgsfaktoren für die Spitzenforschung und Industrie wird zukünftig darin bestehen, jederzeit schnell und standortunabhängig auf weltweit verteilte Daten und IT-Ressourcen zugreifen zu können. Die Grundlage dazu bildet eine hochkomplexe, weltweit vernetzte Informationsinfrastruktur mit differenzierten Zugangs- und Administrationsmechanismen, für die sich allgemein der Begriff "Grid-Computing" etabliert hat. [...] Die ursprüngliche Motivation für Grid-Computing war, daß Rechenkapazität in gleicher Weise universell und transparent zur Verfügung gestellt werden soll, wie dies heute für elektrische Energie der Fall ist."[3]

Damit sehen wir auch, woher der Begriff "Grid" in Grid-Computing stammt, nämlich von der englischen Bezeichnung für das Stromnetz, welches in der englischen Sprache "Power Grid" heißt. Und in der Tat liegt eine deutliche Analogie zwischen Power-Grids und einem Computer-Grid vor. Denn genausowenig, wie man gewöhnlich darüber nachdenkt, wie der elektrische Strom eigentlich ins Haus kommt, so wenig soll man sich demnächst Gedanken darüber zu machen brauchen, wie im Moment benötigte Rechenleistung (oder allgemeiner: benötigte Ressourcen) für den eigenen Computer zur Verfügung gestellt werden können. Diese Form Sorglosigkeit möchte Grid Computing erreichen - Ressourcen sollen einmal überall und permanent zugänglich sein.

Ein in unserem Leben nicht mehr wegzudenkender Teil ist das Internet. Es übermittelt uns Informationen, ohne daß wir uns im Detail mit Organisation und Topologie des Netzes beschäftigen müßten. Grid-Computing möchte das Internet an dieser Stelle erweitern. Es versucht, die Beschränkung des Internets auf reine Informationsübertragung aufzuheben und stattdessen Rechenleistung, Speicherkapazitäten oder gar ganze Applikationen für die Netz-Teilnehmer bereitzustellen.

### Mögliche Einsatzgebiete für das Grid-Computing

Gemäß [9] eignet sich das Grid[1] für folgende Aufgaben:

[1] Gemeint ist an dieser Stelle das ZetaGRID, das wir uns im weiteren Verlauf dieser Seminararbeit noch genauer ansehen werden.

- Mathematische Berechnungen
  - Faktorisierung von Zahlen
  - Berechnung von Nullstellen komplexer Funktionen
- Berechnung finiter Elemente
- Risikobewertung im Bankwesen
- Chip-Design in der Industrie
- Crashsimulationen
- Ölfeldsuche
- Wettervorhersage

Diese Liste erhebt natürlich keinesfalls den Anspruch auf Vollständigkeit. Vielmehr ist eine ganze Reihe weiterer Anwendungsbereiche denkbar.

### Zu dieser Seminararbeit

Diese Seminararbeit soll einen Einblick in das Gebiet Grid-Computing geben. Dazu stellen wir in Kapitel 2 zunächst die wesentliche Funktionsweise sowie verschiedene Anwendungstypen eines Grids vor. Kapitel 3 gibt sodann einen Überblick über den technischen Aufbau eines Computational Grids. Kapitel 4 veranschaulicht die Theorie durch Beispiele existierender Grids. Im fünften und letzten Kapitel blicken wir dann über Tellerrand des Grids hinaus und ziehen Parallelen zu anderen Technologien in der Computerbranche, bevor mit der Diskussion einiger Fragen und Probleme des Grid-Computing diese Seminararbeit abschließen.

Das Material, aus welchem diese Arbeit letztlich entstanden ist, ergibt sich aus der im Anhang aufgeführten Literaturliste. Zitate aus den dort genannten Quellen werden an den entsprechenden Stellen jeweils durch eckige Klammern gekennzeichnet.

Norbert Heidenbluth,
im Juli 2003

# Kapitel 2

# Anwendungstypen des Grids

> "Grid-Computing, most simply stated, is distributed computing taken to the next evolutionary level. The goal is to create the illusion of a simple yet large and powerful self managing virtual computer out of a large collection of connected heterogenous systems sharing various combinations of resources."[4]

Gemäß dieser Definition soll ein Grid verstanden werden als ein sich selbst organisierender "Super-Computer", bestehend aus einer Vielzahl unterschiedlichster Systeme.

Ein wichtiger Aspekt hierbei ist das Attribut "unterschiedlich". In der Terminologie unseres Themas bedeutet dies, daß es sich im allgemeinen um *heterogene* Netze handeln wird und handeln soll. Dabei meint heterogen, daß die teilnehmenden Rechner beispielsweise

heterogene Netze

- unterschiedliche Prozessortypen und
- unterschiedliche Betriebssysteme

verwenden, sowie über

- unterschiedliche angeschlossene Geräte,
- unterschiedliche Kapazitäten und
- unterschiedliche Ausstattungen

verfügen. Jeder der oben genannten Punkte kann - aus dem Blickwinkel des Grid-Computings betrachtet - ein Kriterium darstellen, eine Rechenaufgabe an einen ganz bestimmten Rechner zu verweisen. So zum Beispiel, weil eine Software verwendet werden soll, die an ein bestimmtes Betriebssystem gebunden ist[1]. Einen anderen Grund dafür, bestimmten Rechnern Aufgaben zuzuteilen oder gerade eben *nicht* nicht zuzuteilen, stellen die sogenannten *Policies* dar. Dabei bezeichnen Policies bestimmte Bedingungen, die an eine Ressource geknüpft sein können. So könnte eine Bedingung zum Beispiel lauten, daß ein Rechner nur für medizinische Zwecke benutzt werden darf, oder gerade nicht für militärische Zwecke [4].

Policies

Ein weiterer Wunsch für Grid-Computing-Netze ist *Dynamik*. Dies meint, daß in einem Grid neue Quellen (d.h. Ressourcen) im laufenden Betrieb hinzugefügt oder entfernt werden können. Damit dies möglich wird, sollte die Grid-Architektur auf sogenannten *Agenten* basieren. Agenten sind hier eigenständige Einheiten (im allgemeinen Programme), die für einen Benutzer das Netzwerk nach Aufgaben oder Ressourcen durchsuchen.

Dynamik

Agenten

Welche Anwendungsgebiete sind aber nun für eine derartige Rechenleistung vorstellbar?

Obwohl Grids derzeit nur als Pilotprojekte im Einsatz sind, kann man bereits heute schon fünf Anwendungstypen des Grid-Computings unterscheiden:

- Verteiltes Supercomputing
- Data-Grids, d.h.
  - Computing mit hohem Datendurchsatz
  - Computing von datenintensiven Anwendungen
- Computing bei Bedarf
- Kollaborative Computerarbeit

In diesem Kapitel werden wir zunächst zwei Begriffsdefinitionen vornehmen, um uns sodann einige dieser Anwendungstypen ein wenig näher anschauen.

[1] Auf der anderen Seite ist man jedoch bestrebt, einheitliche Schnittstellen zu finden, welche Grid-Computing über Hard- und Softwaregrenzen hinweg ermöglichen.

## 2.1 Begriffsdefinitionen

Im Rahmen dieser Seminararbeit werden wir einige Begriffe immer wieder benutzen. Zwei dieser grundlegenden Begriffe sind die Wörter *application* und *job*. Sie sollen in diesem Abschnitt eingeführt werden.

### Anwendungen (applications)

Die innerhalb des Grids zur Verfügung stehenden Ressourcen müssen auf eine bestimmte Weise ansprechbar sein. Typischerweise geschieht der Zugriff auf diese Ressourcen durch eine *Anwendung* (engl. "application"). Wir wollen unter diesem Begriff eine Arbeitsanweisung verstehen, die unmittelbar auf einer Ressource, also "in der letzten Instanz" innerhalb unseres Grids, ausgeführt wird.

### Jobs und Subjobs

Anwendungen können aus kleineren Anwendungen zusammengesetzt sein. Die kleineren Anwendungen bezeichnet man dann als *Jobs*. Jobs wiederum können ihrerseits aus kleineren Jobs bestehen, sogenannten *Subjobs*.

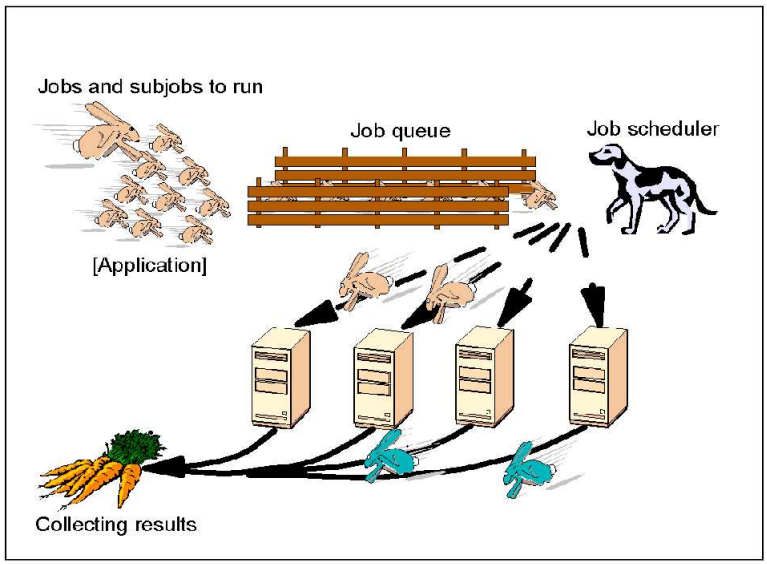
Jobs werden innerhalb des Grids auf einem eigens für sie passenden Ort ausgeführt. Ihre Aufgabe kann darin bestehen, etwas zu berechnen, Daten zu suchen, Systembefehle auszuführen oder angeschlossene Geräte (z.B. Maschinen) zu steuern. Ebenso können mehrere Jobs voneinander abhängig sein und erst weiterarbeiten, wenn ein anderer Job Ergebnisse zurückliefert.

Wir sehen, daß Anwendungen, Jobs und Subjobs eine Hierarchie bilden. In jedem Fall muß an der obersten Stelle der Hierarchie ein Mechanismus existieren, der die Ergebnisse der einzelnen Jobs und Subjobs aufsammelt und somit das eigentliche gewünschte Ergebnis produziert. Das Schaubild 2.1 auf der nächsten Seite soll das soeben beschriebene Vorgehen einmal illustrieren.

## 2.2 Rechen-Grids

Unter dem Begriff Rechen-Grid (engl. *computational grid*) versteht man die Zusammenfassung der Rechenleistung mehrerer Rechner, wie man es bereits von Clustern kennt. Für diese Form des Grid-Computings ist ebenfalls der Begriff *Verteiltes Supercomputing* bekannt.

Abbildung 2.1: Anwendungen, Jobs und Subjobs (aus [4])




### 2.2.1 Ausnutzen nicht verwendeter Ressourcen

Der einfachste Weg, sich diesem Anwendungstyp für das Grid-Computing zu nähern, führt uns wieder zurück zu unserem Sportwagen-Beispiel aus der Einführung. Nach statistischen Erhebungen sind Desktop-PCs in den meisten Organisationen weniger als 5% des Tages wirklich beschäftigt[4]. Teilweise trifft dies ebenso auf große Server zu. Und selbstverständlich darf man hier nicht die privaten PCs vergessen, die den größten Teil des Tages ausgeschaltet und damit unproduktiv auf dem Schreibtisch stehen.

Grid-Computing in der in diesem Abschnitt besprochenen Anwendungsform möchte auf diese ungenutzten Kapazitäten zugreifen. So kann es beispielsweise sein, daß ein Großrechner mit einer Aufgabe beschäftigt ist, welche ihn außergewöhnlich auslastet. Stellen wir uns vor, daß er nun eine wichtige weitere Aufgabe bekommen soll, für diese jedoch keine weiteren Kapazitäten mehr bereit hält. Mittels des Grid-Computings können wir diese neue Aufgabe einem weniger oder gar nicht beschäftigten Rechner innerhalb unseres Grids übertragen. Dabei sind folgende Vorbedingungen an den entfernten Rechner zu beachten:

- Die Aufgabe muß ohne unverhältnismäßigen Aufwand auf dem entfernten Rechner zu erledigen sein.

- Der Rechner im Grid, der die Aufgabe erhalten soll, muß über die für diese Aufgabe benötigte Hardware, Software und Ressource(n) verfügen.

### 2.2.2 Parallele CPU-Kapazitäten

Ein weiteres interessantes Anwendungsgebiet für Grid-Computing in verteilten Systemen liegt in der Aufteilung von rechenintensiven Anwendungen auf mehrere Computer. Man kann sich vorstellen, daß eine solche Anwendung in viele kleine Unteranwendungen (Subjobs) zerfällt, welche auf verschiedene Computer aufgeteilt werden. Die teilnehmenden Computer rechnen also parallel jeweils einen kleinen Teil der großen Aufgabe. Hierzu sind sie per Internet miteinander verbunden. Deshalb nennt man diese Form des Grid Computings auch *Internet Computing*.

Das Problem in diesem Anwendungsgebiet liegt in der *Skalierbarkeit* der Aufgabe. Skalierbarkeit bedeutet dabei, inwieweit die Unteranwendungen untereinander abhängig sind und beispielsweise miteinander kommunizieren müssen. Der Idealfall der *perfekten Skalierbarkeit* liegt vor, wenn die Unteranwendungen völlig unabhängig voneinander sind und führt dazu, daß eine Anwendung 10 mal schneller fertig ist, wenn 10 Rechner teilnehmen. Skalierbarkeit

Perfekte Skalierbarkeit kann aber nicht immer erreicht werden. Grenzen für perfekte Skalierbarkeit liegen zum Beispiel vor, wenn

- Algorithmen nur in eine begrenzte Anzahl unabhängiger (Sub-)jobs aufgespalten werden können
- ein gemeinsamer Zugriff mehrerer (Sub-)jobs auf eine einzelne Ressource (z.B. Datei oder Datenbank) nötig ist. An dieser Stelle kann dann ein “Stau” entstehen.

### 2.2.3 Geeignete Aufgaben für verteiltes Supercomputing

Nicht alle Aufgaben können für verteiltes Supercomputing gleichermaßen geeignet sein. In [9] finden sich folgende Anforderungen an eine geeignete Aufgabe:

**Muß-Bedingungen**

- Rechenintensive Aufgabe, die einzelne Rechner überfordert
- Aufgabe muß gut unterteilbar (parallelisierbar) sein
- Keine Benutzerinteraktionen notwendig
- Möglichkeit der Speicherung von Zwischenergebnissen und Wiederaufnahme bei der Verwendung von Bildschirmschonern (engl: "*checkpointing*").
- Der genaue Datendurchsatz ist schwer vorhersagbar, daher muß entweder genug Kapazität vorhanden sein oder die Aufgabe sollte nicht zeitkritisch sein.

**Kann-Bedingungen**

- Möglichst wenig bzw. kleinere Datenzugriffe nötig
- Möglichst in sich geschlossene Aufgabe (kein "Riesenprogramm", kein komplex zu konfigurierendes Programm)

## 2.3 Data-Grids

### 2.3.1 Temporärer vs. permanenter Speicherplatz

Die zweithäufigste Anwendungstyp des Grid Computings stellen Daten-Grids dar. Unter einem Daten-Grid (engl. "Data-Grid") versteht man ein Grid, in dem jeder Rechner Speicherplatz für die Daten anderer Grid-Teilnehmer zur Verfügung stellt. Dies kann permanenter oder auch nur temporärer Speicherplatz sein. Letzteren kann man verwenden, um einen schnelleren Datenzugriff zu ermöglichen. Der Nachteil in diesem Fall ist natürlich, daß dieser Speicherplatz nur "flüchtig" ist. Als Beispiel sei an dieser Stelle der RAM-Speicher genannt. Permanenter Speicherplatz hingegen dient dazu, um Speicherkapazitäten und Performance zu erhöhen, ist aber auch im Hinblick auf die Sicherheit der Daten vor Datenverlust interessant.

#### Sicherheit vor Datenverlust

Wie kann das Grid vor dem Verlust von Daten schützen? Zum Beispiel durch mehrfache (redundante) Datenhaltung: Daten können dupliziert werden und auf mehreren Rechnern abliegen, um beispielsweise als Backup zu dienen.

#### Verbesserung der Perfomance

Eine weitere Frage, die sich in diesem Abschnitt ergibt, ist die, was das Grid zur Verbesserung der Performance beitragen kann? Auch hier liegt die Antwort in der redundanten Datenablage. Man spricht in diesem Zusammenhang auch von *replication*. Dabei besteht nach [11] der Hauptzweck der *replication* darin, bessere Datenübertragungsgeschwindigkeiten dadurch zu erreichen, daß die Daten von lokal "gecachten" Datenbanken bezogen werden, anstatt jeweils einzeln über das gesamte Netzwerk übertragen werden zu müssen. Die Daten können von der jeweils naheliegendsten Quelle anstatt von einem zentralen Server übermittelt werden. replication

Deshalb bedienen sich Grid-Systeme oftmals sogenannter *Mountable Network File Systems*, also solcher Dateisysteme, die das Ein- und Aushängen von Verzeichnissen via Netzwerk ermöglichen. Beispiele hierfür sind Andrew File System (AFS), Network File System (NFS), Distributed File System (DFS) oder das General Parallel File System (GPFS).

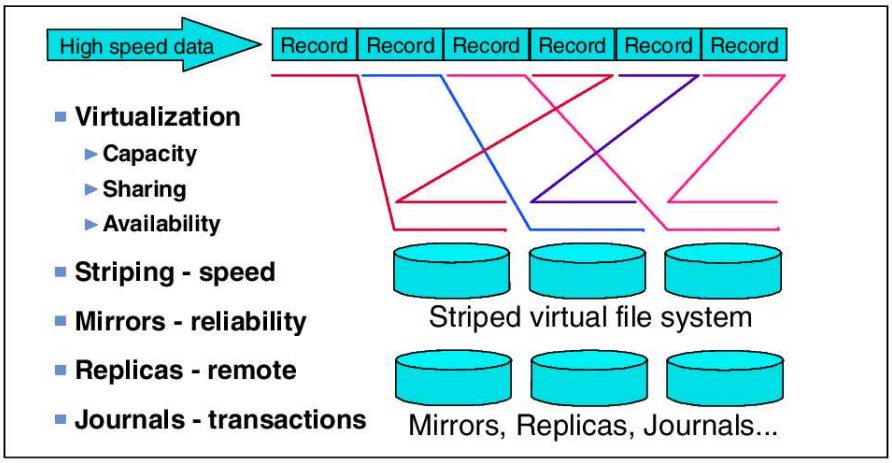
### 2.3.2 Data Striping

Einen weiteren interessanten Ansatzpunkt, um eine Verbesserung der Perfomance beim Datenzugriff zu erreichen, stellt das *Data Striping* dar. Hierunter versteht man eine Technik, die den Effekt eines schnelleren Datenzugriffs simuliert. Diese funktioniert wie folgt:

Die Daten liegen auf mehreren physischen Geräten zugleich ab. Wenn der Zugriff auf diese Daten auf eine im voraus bekannte Weise geschieht (z.B. sequentiell), so werden die Daten nicht von nur einem Gerät gelesen[2], sondern nacheinander zeitversetzt von allen diesen Geräten. Auf diese Art entsteht der Eindruck, als lese man von einem besonders schnellen Gerät. Tatsächlich aber überlappen sich die Zugriffe auf die verschiedenen Geräte und führen so zu diesem Eindruck. Abbildung 2.2 auf der nächsten Seite versucht, dieses Verfahren zu veranschaulichen.

[2]Das Gleiche gilt natürlich auch für das Schreiben von Daten.

Abbildung 2.2: Data Striping (aus [4])




## 2.4 Kollaborative Computerarbeit

### 2.4.1 Virtuelle Ressourcen und virtuelle Organisationen

Das Grid kann heterogene und geographisch weit entfernte Ressourcen zu einer sogenannten *virtuellen Organisation*(VO) zusammenführen. Ein Benutzer des Grids kann mehreren virtuellen Organisationen angehören, welche wiederum zusammen ein noch größeres Grid darstellen. Virtuelle Organisationen zeichnen sich nach [12] insbesondere dadurch aus, daß sie einen

virtuelle Organisation

- in hohem Maße kontrollierten,
- von Ressourcen-Anbietern und -Konsumenten klar definierten
- in sich geschlossenen[3]

Austausch von Ressourcen ermöglichen. Dabei bezieht sich dieser Ressourcenaustausch nicht bloß auf Dateien, sondern ebenfalls auf direkten Zugriff auf entfernte Rechner, Software, Daten und Hardware, die von für eine bestimmte Problemlösung zusammenarbeitenden Computern benötigt werden.

[3]Dies soll bedeuten, daß festgelegt ist, wer am Austausch teilnehmen darf und wer nicht.

Virtuellen Organisationen kommt nach [12] in der Zukunft noch eine besondere Bedeutung zu:

> "It is our belief that VOs have the potential to change dramatically the way we use computers to solve problems, much as the web has changed how we exchange information."

Innerhalb einer virtuellen Organisation können Daten in Form von Dateien oder Datenbanken für alle Mitglieder der Organisation zugänglich gemacht und damit *geteilt* werden. Wenn jedes Mitglied Zugriff auf seine Daten ermöglicht, so ergibt auch dies ein *Data-Grid*, also gewissermaßen eine sehr große Datenbank, von der alle Teilnehmer profitieren können. Zur näheren Erläuterung des Data-Grids im Allgemeinen: siehe auch den vorigen Abschnitt.

Im Zusammenhang mit virtuellen Organisationen ist ein Data-Grid ist aber nicht nur auf Daten beschränkt. So können hier auch andere Ressourcen geteilt werden, wie

**Hardware** (z.B. Bandbreite beim Internet-Zugang): teilt man den Zugang auf das Internet auf mehrere (unabhängige) Rechner auf, so erhält man ein Vielfaches der Kapazität, wenn man beispielsweise Data-Mining-Suchmaschinen verwendet.[4]

**Software / Lizenzen:** wenn auf einzelnen Rechnern teure Software installiert ist, so können andere Rechner diese Software benutzen, um die Lizenz auf diese Weise besser auszunutzen.

**Service:** teure Spezialgeräte (z.B. ein Elektronenmikroskop) können über das Grid benutzt werden. Man sendet die zu untersuchende Probe an die Einrichtung, in der das Mikroskop steht, und steuert das Mikroskop vom entfernten Computer aus.

Wichtig bei den beiden letztgenannten Beispielen ist natürlich, daß vorher zeitliche Absprachen getroffen werden. Der nächste Abschnitt befasst sich daher mit der Problematik der Zugriffs-Organisation innerhalb eines Grids.

### 2.4.2 Reservation, Scheduling und Scavenging

Die Begriffe *reservation*, *scheduling* und *scavenging* werden im Zusammenhang mit Grid Computing dazu verwendet, anstehende Jobs zu organisieren.

[4] Andere Beispiele wären hier besondere Drucker, schnelle CD/DVD-Brenner etc.

### Reservation

Der einfachste Fall der Organisation ist hier das Reservieren (engl. "reservation") von benötigten Grid-Ressourcen. Die im letzten Abschnitt erwähnte zeitliche Absprache, um zum Beispiel das Elektronenmikroskop zur Verfügung stehen zu haben, fällt unter diesen Begriff.

Allerdings wird beim Grid Computing diese zeitliche Absprache nicht unbedingt unter Menschen, sondern unter Maschinen vorgenommen. Es existiert innerhalb des Grids ein sogenanntes *reservation system*, welches eine Ressource für ein bestimmtes Datum und einen genau definierten Zeitraum für einen bestimmten Rechner zur Verfügung stellt. Dazu muß das Reservierungssystem die Berechtigung haben, andere Jobs, die zu dem reservierten Zeitpunkt auf die Ressource zugreifen, anzuhalten oder abzubrechen. Selbstverständlich muß es auch dafür Sorge tragen, daß es keine Überschneidungen bei der Reservierung gibt.

### Scheduling

Ein weiterer Organisationsmechanismus ist das *Scheduling*. Eine organisierende Instanz des Grids, der sogenannte *job scheduler* sucht automatisch die passendste Maschine auf dem Grid aus, um einen anstehenden Job an diese zu übertragen. Der Scheduler kennen die augenblickliche Verfügbarkeit aller Grid-Ressourcen und kann daher Jobs optimal zuweisen. Um die Optimalität der ausgewählten Maschinen zu gewährleisten, bedarf es jedoch ausgeklügelter und sehr schwer zu bestimmender Optimisierungs-Algorithmen.

Weitere Aufgaben für den Scheduler können darin bestehen, die erfolgreiche Beendigung von vermittelten Jobs zu überwachen. Im Falle von nicht erfolgreichen Ausführungen sollte ein guter Scheduler den Job ein weiteres Mal vermitteln (eventuell an eine andere Maschine). Hingegen sollten Jobs, die in einer Endlos-Schleife enden, nicht erneut vermittelt werden, da dasselbe (unerwünschte) Ergebnis produziert würde.

Schedulers können hierarchisch organisiert werden. Auf diese Weise werden Jobs zunächst mehrere Scheduler-Ebenen passieren, bevor sie schließlich (vom Scheduler der untersten Ebene) einer individuellen Maschine zugewiesen werden.

### Scavenging

Eine dritte Möglichkeit der Organisation bietet ein *scavenging grid system*. Innerhalb eines solchen Grids berichten die teilnehmenden Maschinen ihren

Status an eine zentrale Verwaltungsinstanz (*grid management node*). Insbesondere melden sie, wenn sie gerade nichts zu tun haben. Die Verwaltungsinstanz kann dann gezielt den nächsten Job an die untätige Maschine vermitteln, vorausgesetzt, daß sie den Anforderungen des Jobs genügt.

Um beim Scavenging nicht in Konflikt mit Prozessen zu gelangen, die lokal auf den Maschinen laufen sollen und nicht mit Grid Computing in Verbindung stehen, werden Scavenging-Systeme meist dergestalt implementiert, daß lokale Prozesse Priorität besitzen und dazu führen, daß ein laufender Grid-Job angehalten wird, um die Kapazitäten für den lokalen Prozeß freizugeben. In diesem Umstand liegt auch der größte Nachteil des Scavenging, denn es ist nicht vorhersehbar, wie lange ein Grid-Prozeß braucht, um auf einem solchen System beendet zu werden. Umgehen kann man dieses Problem natürlich, indem man ein Scavenging System ausschließlich für Grid-Jobs bereitsstellt.

### 2.4.3 Ausgleich von Resourcen

Das Grid kann dafür sorgen, daß virtuelle Ressourcen stets weitgehend gleichmäßig ausgelastet sind. Hierzu verteilt das Grid Aufgaben auf Rechner, die derzeit wenig oder gar nicht beansprucht werden. Dieses Verfahren kann von enormer Bedeutung sein, wenn einzelne Rechner besonders stark ausgelastet sind. Grundsätzlich bestehen zwei Möglichkeiten, einer weiteren Belastung zu entgehen:

- Der außergewöhnlich belastete Rechner verteilt Aufgaben auf weniger ausgelastete Rechner
- Aufgaben mit nur geringer Priorität werden vorübergehend angehalten (suspendiert), um Kapazitäten für die höher priorisierte Aufgabe zu schaffen.

Diese Eigenschaft des Grids kann sich für termingebundene rechenintensive Aufgaben als sehr nützlich erweisen. Denn wenn der Termin der Fertigstellung (*Deadline*) immer näher rückt, so kann man auf diese Weise die Aufgabe in einzelne Unteraufgaben teilen und durch die einzelnen Rechner des Grids erledigen lassen. So ist eine termingerechte Fertigstellung vielleicht noch zu erreichen.

### 2.4.4 Betriebssicherheit und Zuverlässigkeit

#### Ausfallsicherheit

Herkömmliche High-End-Computersysteme verwenden derzeit noch teure Hardwarevorkehrungen, um eine gewisse Ausfallsicherheit zu gewährleisten. Diese Vorkehrungen können beispielsweise sein

- redundante Schaltkreise
- doppelt vorhandene Prozessoren, von denen einer bei laufendem Betrieb ausgetauscht werden kann, während der andere weiterarbeitet
- Wiederherstellungsmechanismen, die nach einem Hardwarefehler greifen
- Doppelt vorhandene Netzteile und Kühlsysteme
- Spezielle, gegen Ausfälle abgesicherte Stromkreisläufe

Diesen Vorkehrungen ist gemein, daß sie hohe Kosten verursachen, da teure Hardwarekomponenten teilweise doppelt vorhanden sind. Anstelle der teuren Hardware wird es durch die Anwendung von Grid Computing demnächst Software sein, die kritische Situtationen meistert. Fällt beispielsweise ein Rechner innerhalb des Grids wegen Stromausfalls aus, können die anderen jedoch unbeschadet weiterarbeiten. Grid Management Software kann in diesem Fall die auf dem ausgefallenen Rechner laufenden Jobs erneut an einen anderen Rechner vergeben.

Dasselbe gilt für die Ergebnis-Kontrolle. Wichtige Aufgaben können auf mehreren Rechnern gelöst und ihre Ergebnisse sodann auf Übereinstimmung (Konsistenz) überprüft werden. Auf diese Weise registriert man Rechenfehler, Datenkorruption und sogar Einbrüche in das System.

#### Autonomic Computing

Eine weitere Absicherung von Grid Systemen stellt das sogenannte *Autonomic Computing* dar. Darunter versteht man eine Software, die - soweit möglich - selbständig Probleme innerhalb des Grids behebt, vielleicht sogar bevor ein Administrator diese bemerkt hat.

Auf jeden Fall wird Grid Computing in der Zukunft in der Lage sein, die derzeit noch von Hardware geleistete Ausfallsicherheit durch Software zu erreichen.

# Kapitel 3

# Aufbau und Anwendung eines Grids

In diesem Kapitel beleuchten wir die technische Seite des Grids. Im ersten Abschnitt werden wir ein Grid stufenweise auf- bzw. ausbauen, Abschnitt zwei erläutert uns dann einige Software-Komponenten, die innerhalb eines Grids interagieren, bevor wir uns im dritten und letzten Abschnitt die im Grid verwendeten Protkoll-Schichten anschauen.

Die Ausführungen in diesem Kapitel fassen im Wesentlichen die Ausführungen aus [4] und [12] zusammen. Für detaillierte Erläuterungen sei auf diese Quellen verwiesen.

## 3.1 Der stufenweise Aufbau eines Grids

### Stufe 1: Cluster

Die einfachste Form, ein Computational Grid aufzubauen, besteht darin, einige (wenige) Rechner lokal miteinander zu vernetzen. Dabei sind die beteiligten Rechner typischerweise homogen, daß heißt, sie sind vom gleichen Hardwaretyp und verwenden das gleiche Betriebssystem.

Der Sinn dieses noch recht primitiven Grids kann es sein, Grid Software zu testen. Weil sich die Rechner in unmittelbarer Nachbarschaft befinden (zum Beispiel in der selben Abteilung), braucht man sich um Sicherheit und Policies (vgl. Kapitel 2) nicht besonders zu kümmern.

Diese Form des Computational Grids[1] nennt man *Cluster*.

[1] falls es überhaupt schon eines ist...

### Stufe 2: Intragrid

Im Unterschied zum Cluster in Stufe 1 wollen wir nun unser Netzwerk aus heterogenen Rechnern aufbauen. Auf diese Art und Weise haben wir mehrere Arten von Ressourcen zur Verfügung. Ebenso können bereits einfache Scheduling-Mechanismen implementiert werden. Der Austausch von Dateien ist via Network File Systems möglich.

Ein Anwendungsbeispiel für diesen Grid-Aufbau ist zum Beispiel die Vernetzung von Rechnern mehrerer Abteilungen innerhalb eines Unternehmens. Auf diese Weise kommt diese Stufe auch zu ihrem Namen: *Intragrid*.

### Stufe 3: Intergrid

Wenn das Intragrid expandiert und immer mehr Abteilungen verbindet, die auch geographisch weit auseinander liegen können, ist es an der Zeit, sich mehr Gedanken über die Sicherheit, über Policies und über Scheduling zu machen. Treten nun auch noch weitere Unternehmen dem Grid bei und befinden sich die teilnehmenden Rechner weltweit verstreut, so haben wir die Stufe 3 erreicht und betrachten unser Grid als *Intergrid*.

## 3.2 Grid Software Komponenten

In einem Grid interagieren verschiedene Software-Komponenten miteinander, die man unterschiedlichen Klassen zuordnen kann. Sehen wir uns nachfolgend nun einige dieser Komponenten etwas genauer an.

### 3.2.1 Management Komponenten

Jedes Grid-System muß über mehrere Management Komponenten verfügen. Diese haben folgende administrative Aufgaben:

- Koordination der vorhandenen Ressourcen
- Ermittlung von Informationen über Auslastung und Verfügbarkeit von Ressourcen, Verwendung diesen Informationen für
  - Durchführung des Scheduling
  - Erkennen von Ausfällen einzelner Ressourcen

  - Alarmierung des Administrators
  - Statistische Erhebungen

- Automatische Wiederherstellung der Integrität bei Störungen (vgl. *Autonomic Computing* in Kapitel 2)

### 3.2.2 Donor Software Komponenten

Donor Software Komponenten werden auf den Rechnern installiert, die dem Grid eine Ressource zur Verfügung stellen ("spenden"). Sie dienen der Ausführung von Befehlen oder Programmen auf dem spendenden Rechner. Die Donor Software kommuniziert mit dem Management System, von welcher es den auszuführenden Job bekommt.

Eine weitere Aufgabe der Donor Software besteht darin, den momentanen Auslastungsstatus des Rechners an die Management-Software zu berichten.

### 3.2.3 Submission Software Komponenten

Diese Software Komponente findet man nicht zwingend in jedem Grid. Sie dient dazu, die Übermittlung von Jobs in das Grid zu organisieren.

### 3.2.4 Scheduling Software Komponenten

Scheduling Software Komponenten kümmern sich um alles, was in Zusammenhang mit Scheduling steht. Dieses wurde bereits auf Seite 16 erläutert.

### 3.2.5 Communication Software Komponenten

Unter Umständen müssen einzelne Jobs oder Subjobs miteinander kommunizieren. Die Communication Software dient dazu, daß sich die einzelnen Subjobs untereinander "finden" , einen Kommunikationskanal aufbauen und die erforderlichen Informationen austauschen können.

Für diese Art der Kommunikation wird oftmals das "Open Standard Message Passing Interface" (MPI) verwendet.

## 3.3 Grid-Architektur

Wie zuvor schon erwähnt, soll das Grid heterogenen Rechnern die Zusammenarbeit ermöglichen. Insoweit ist die Zusammenarbeit unterschiedlichster Systeme (engl.: *interoperability*) der zentrale Punkt beim Aufbau eines Grids. interoperability

Um die gewünschte Zusammenarbeit unterschiedlicher (vernetzter) Systeme zu erreichen, baut man das Grid auf einer Protokollarchitektur auf, in der Protokolle die grundlegenden Mechanismen des Verbindungsaufbaus und den Daten- bzw. Ressourcenaustausch übernehmen. Eine auf Standard-Protokollen aufgebaute Architektur vereinfacht

- die Erweiterbarkeit,
- die Zusammenarbeit,
- die Portabilität,
- und das Code-Sharing.

Entwicklungsumgebungen, die sich an dieser Technologie und Architektur orientieren, nennt man "*middleware*". In [12] wird der Begriff middleware wie folgt definiert: "[Middleware are] *the services needed to support a common set of applications in a distributed network environment*". middleware

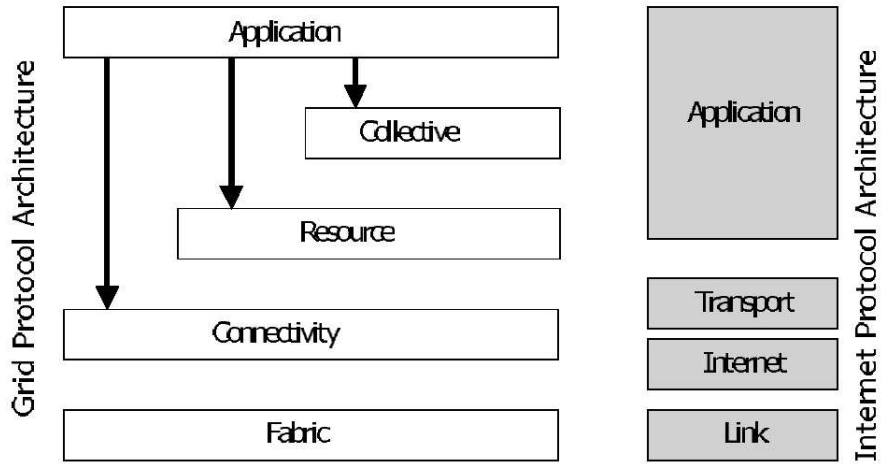
Die Protokoll-Architektur des Grid Computings lehnt sich dabei stark an die Internetprotokolle an. Abbildung 3.1 auf der nächsten Seite stellt die beiden Protokollfamilien einmal gegenüber. Sehen wir uns im Folgenden die Aufgaben der einzelnen Schichten an:

### 3.3.1 Fabric Layer

Die Fabric-Schicht liefert die Ressourcen, die für einen gridweiten Zugriff zur Verfügung gestellt werden sollen. Beispiele hierfür sind unter anderem

- Rechenkapazitäten
- Arbeits- und Datenspeicher
- Netzwerk-Ressourcen
- Sensoren

Abbildung 3.1: Gegenüberstellung Grid- bzw. Internet Protocol Architecture (aus [12])




Eine Ressource kann dabei eine physische oder eine logische Einheit sein, wie zum Beispiel

- Computercluster
- verteilte Dateisysteme

Im letzteren Fall liefert die Fabric-Schicht diejenigen Protokolle, die nötig sind, um außerhalb der logischen Einheit zu kommunizieren. Innerhalb der Einheit sind eigene Protokolle notwendig, die allerdings nicht das Grid betreffen.

Allgemein kann man sagen, daß die Fabric-Schicht "lokale, ressourcenspezifische Operationen implementiert, die für einzelne (physische oder logische) Ressourcen notwendig sind[12]". Zu beachten ist, daß auf dieser Schicht bereits grundlegende Fähigkeiten des Daten- und Ressourcenaustauschen implementiert werden. Je mehr Funktionalität diese Schicht besitzt, desto mehr Möglichkeiten zum Austausch werden eröffnet.

### 3.3.2 Connectivity Layer

Die Connectivity-Schicht definiert die grundlegenden Kommunikations- und Authentifizierungsprotokolle, die für Grid-typische Anwendungen benötigt werden. Diese beiden Protokollen besitzen folgende Eigenschaften:

**Kommunikationsprotokolle** ermöglichen den Austausch von Daten zwischen Ressourcen auf der Fabric-Schicht. Auch wenn bereits Alternativen existieren, werden diese derzeit noch aus den bewährten Protokolle der TCP/IP-Familie abgeleitet.

**Authentifizierungsprotokolle** erweitern Kommunikationsdienste um Sicherheitsmechanismen, die verwendet werden, um Identitäten von Benutzern oder Ressourcen zu überprüfen. Die in Abschnitt 3.6 auf Seite 27 genannten Sicherheitsmechanismen würden an dieser Stelle implementiert.

### 3.3.3 Resource Layer

Die Ressourcen-Schicht dient - vereinfacht gesagt - zur Überwachung und Steuerung einer einzelnen Ressource. Es lassen sich zwei grundlegende Klassen von Protokollen dieser Schicht unterscheiden:

**Information protocols** werden verwendet, um Struktur- und Statusinformationen über eine bestimmte Ressource zu erhalten (z.B. momentane Auslastung, Konfiguration)

**Management protocols** werden verwendet, um den Zugriff auf die Ressource zu kontrollieren.

### 3.3.4 Collective Layer

Die Collective-Schicht ist - in Erweiterung der soeben vorgestellten Ressourcen-Schicht - nicht nur für eine einzelne Ressource verantwortlich, sondern für ein "Kollektiv", also eine ganze Sammlung von Ressourcen.

### 3.3.5 Applications Layer

Die Anwendungsschicht bezieht sich auf die Programme (also die Anwendungen), die ein Benutzer innerhalb des Grids bzw. der virtuellen Organisation verwendet. Diese Schicht sorgt dafür, daß die Anwendung auf die Dienste der anderen Schichten über vordefinierte Protokolle zugreift und somit letztlich zur gewünschten Ressource vordringt.

## 3.4 Das Globus Toolkit

Nicht unerwähnt bleiben sollte in diesem Zusammenhang das Globus-Toolkit. Es stellt eine Entwicklungsumgebung für Grid-Anwendungen zur Verfügung, welche einheitliche Zugangsschnittstellen für unterschiedliche Rechner-Architekturen bietet. Dabei orientiert sich das Toolkit am vorstehend beschriebenen Schichtenmodell.

Das Globus Toolkit ist derzeit so etwas wie ein Standard. Für weitere Informationen hierüber sei an dieser Stelle auf die Homepage des Projektes verwiesen. Sie lautet: `http://www.globus.org`.

## 3.5 Die Verwendung eines Grids

In diesem Abschnitt wollen wir das Grid einmal aus der Perspektive eines Grid-Benutzers betrachten.

### 3.5.1 Schritt 1: Installation der Grid-Software

Der erste Schritt, um in den Genuß des Grid Computings zu gelangen, ist sicherlich die Installation der Grid-Software auf dem Computer des Benutzers. Je nach Ausprägung der Sicherheitsmechanismen innerhalb des Grids kann für die Durchführung der Installation eine Überprüfung erforderlich sein, ob die angegeben persönlichen Daten des neuen Nutzers auch der Wahrheit entsprechen. Eine solche Kontrolle darüber hinaus auch für den eigentlichen Computer (und vielleicht sogar allgemeiner: für die Ressourcen) des Benutzers denkbar.

Während der Installation hat der Benutzer die Wahl, ob er nur auf Grid-Ressourcen zugreifen möchte, oder auch selbst eigene Ressourcen im Grid zur Verfügung stellen möchte. In letzterem Fall nennt man den Benutzer auch *Spender* (engl.: "*donor*"). Spender

### 3.5.2 Schritt 2: Anmeldung am Grid

Um am Grid Computing teilzunehmen, ist aus Sicherheitsgründen üblicherweise eine Anmeldung am Grid erforderlich. Ist die Anmeldung erfolgt, so kann der Benutzer Jobs an das Grid übermitteln und Anfragen an das Grid starten.

### 3.5.3 Schritt 3: Anfragen und Übermittlung von Jobs an das Grid

Ein Beispiel für eine Anfrage an das Grid ist zum Beispiel die Frage, welche Ressourcen derzeit verfügbar sind, und welche Auslastung das Grid derzeit hat. Zum Erstellen dieser Anfragen sind einerseits Befehle denkbar, die in eine Kommandozeile eingegeben werden, andererseits existieren aber ebenfalls häufig graphische Benutzerschnittstellen (GUI), um die Anfragen zu erstellen. Der Vorteil von einer auf Kommandozeilen basierenden Lösung ist, daß der Benutzer hier Skripte schreiben und auf diese Weise eine Automatisierung seiner üblichen Anfragen erreichen kann.

Die Übermittlung (oder besser: die Ausführung) von Jobs zerfällt in drei Teile:

1. Senden des ausführbaren Programms auf den Rechner, der die Aufgabe berechnen soll
2. Ausführung des Jobs auf dem entfernten Rechner
3. Rücksendung der Ergebnisse an den Benutzer

Auch für die Übermittlung von Jobs kann eine kommandozeilenbasierte Grid-Software von Vorteil sein. So zum Beispiel, wenn es sich bei dem auszuführenden Job um die Aufgabe handelt, optimale Parameter-Kombinationen zu bestimmen. Der Benutzer kann zu diesem Zweck auch hier Skripte erstellen, die ein und denselben Job hintereinander mit variierten Parametern ausführen lassen. Auf diese Weise lassen sich auf einfache Weise die besten Parameterkombinationen ermitteln.

### 3.5.4 Schritt 4: Statusüberwachung und Wiederherstellung

Üblicherweise wird dem Benutzer in Grid-Software die Möglichkeit geboten, die Abarbeitung seines Jobs zu verfolgen. Insbesondere kann dies von Interesse sein, wenn ein Job in viele Subjobs aufgespalten wird.

Was geschieht nun aber, wenn einer der Subjobs nicht erfolgreich beendet werden kann, zum Beispiel durch

- Programmierfehler

- Hardwarefehler
- Verbindungsunterbrechung
- Endlosscheife

oder andere Fehler? Die bereits auf Seite 16 erwähnten Scheduler können in diesem Fall möglicherweise die Fehlerursache ermitteln und den Job ein weiteres Mal zur Ausführung bringen lassen - eventuell auf einem anderen Grid-Rechner.

### 3.5.5 Schritt 5: Ergebnisauswertung

Das errechnete Ergebnis des Jobs wird an den Rechner übermittelt, der ursprünglich diesen Job initiiert hat. Im einfachsten Fall wäre damit der Job erledigt. Denkbar sind aber auch Überprüfungen, ob das zurückgelieferte Ergebnis plausibel ist und ob es wirklich von dem Rechner zurückgeliefert wurde, der den Job bekommen hatte. Ebenso ist es bei sensiblen Ergebnissen denkbar, daß der Job mehrfach und unabhängig voneinander an mehrere Rechner vergeben wurden. Die zurückgelieferten Ergebnisse werden nun auf Übereinstimmung verglichen.

## 3.6 Sicherheitsaspekte

Zum Ende dieses Kapitels soll noch kurz die Frage der Sicherheit auf dem Grid beantwortet werden. Im Verlauf unseres Textes wurde immer wieder darauf hingewiesen, daß ein Arbeiten auf dem Grid nur nach erfolgreicher Anmeldung mit UserID und Paßwort möglich sein sollte (*authentication*). In [13] wird empfohlen, Authentifizierungsmethoden mit folgenden Eigenschaften auszustatten: authentication

**Single sign on:** Der Benutzer braucht sich nur einmal anzumelden, um ohne weitere Authentifizierungen auf *alle* die Ressourcen zugreifen zu können, für die er Zugriffsberechtigungen hat.

**Delegation:** Es muß die Möglichkeit gegeben sein, daß ein Programm auf dem Grid mit den Rechten des Users, der es gestartet hat, ausgeführt wird. Somit kann das Programm auf alle die Ressourcen zugreifen, auf die der (ausführende) Benutzer Zugriff hat.

Gibt es darüber hinaus noch die Möglichkeit, daß ein Programm Unterrechte an andere Programme vergibt, so spricht man von *restricted delegation*.

**Kompatibilität:** In heterogenen Systemen wird man auf viele diverse Lösungen treffen, um Ressourcen gegen Angriffe von außen und vor unbefugter Nutzung abzusichern. Da es jedem Benutzer freigestellt ist und vor allem auch freigestellt bleiben soll, wie er seine Ressourcen schützt, muß stets gewährleistet sein, daß der Sicherheitsmechanismus des Grids zu den Mechanismen der Ressourcen kompatibel ist.

**User-based trust relationships:** Möchte ein Benutzer auf zwei verschiedene Ressourcen gleichzeitig Zugriff haben und sie zur gleichen Zeit verwenden, dann sollte dies möglich sein, ohne daß sich diese beiden Ressourcen untereinander sicherheitsbedingt verständigen brauchen. Der Sicherheitsmechanismus des Grids sollte also erkennen, daß der Benutzer für beide Ressourcen zugangsberechtigt ist und hieraus bereits das Zusammenarbeiten der beiden Ressourcen ermöglichen, ohne ihre Sicherheitsdienste in Anspruch nehmen zu müssen.

# Kapitel 4

# Beispiele für Grids

In diesem Kapitel werden wir uns einige Beispiele für derzeit existierende Grids anschauen. Das erste Beispiel ist ein seit einiger Zeit erfolgreiches Projekt aus dem Bereich Internet-Computing und repräsentiert ein Computational Grid. Auch das zweite Beispiel demonstriert uns ein Computational Grid, allerdings stellt dieses ein Produkt dar, das auf dem Markt vertrieben werden soll. Das dritte Beispiel befindet sich momentan im Aufbau und beschreibt ein Data-Grid.

## 4.1 SETI@home

### 4.1.1 Ursprung des Projekts

Das erste Grid, das wir uns als Beispiel für ein Computational Grid ansehen, wurde von der Universität Berkeley in Kalifornien (UCB) ins Leben gerufen. Sein Name ist "SETI@home". Dabei ist SETI die Abkürzung für "Search for Extraterrestrial Intelligence".

SETI steht in Verbindung mit dem größten Radioteleskop der Welt, dem *Arecibo Radio Observatory* auf Puerto Rico in der Karibik. Dieses sucht Nacht für Nacht den Himmel nach ungewöhnlichen Radiosignalen ab, in der Hoffnung, hier Signale von Außerirdischen zu erhalten. Auf diese Weise wird täglich eine solche Datenflut aufgezeichnet, die in der Vergangenheit selbst mit Großrechnern nicht zu verarbeiten war.

An dieser Stelle gründete die UCB das Projekt "SETI@home", welches heute[1] mit rund 4,5 Millionen Teilnehmern zu den größten verteilten Projekten der

[1] Stand dieser Aussage: Juni 2003

Tabelle 4.1: Nutzerstatistik bei SETI@home, Stand: 20.06.2003, aus [8]

| | Gesamt | Letzte 24 Stunden |
|---|---:|---:|
| Nutzer | 4.547.489 | 1.703 |
| Empfangene Resultate | 924.936.676 | 1.587.967 |
| CPU-Zeit gesamt | 1.504.538,170 Jahre | 1.713,892 Jahre |
| Floating Point Operations | 3,191338e+21 | 6,193071e+18 |
| Durchschn. Rechenzeit pro work unit | 14 Std. 14 Min. 57.6 Sek. | 9 Std. 27 Min. 16.7 Sek. |

Welt zählt. Jeder Teilnehmer installiert sich auf seinem Rechner zu Hause den SETI-Client, ein kleines Programm, das sich von einem Server Arbeitsaufträge (Auswertung der vom Radioteleskop aufgezeichneten Daten) herunterlädt, beim Teilnehmer die eigentliche Auswertung der Daten durchführt und das Ergebnis an den Server zurückliefert. Diesen Client gibt es für sämtliche Betriebssysteme und Prozessor-Architekturen.

Damit der Teilnehmer die Rechenleistung seines Computers uneingeschränkt zur Verfügung hat, arbeitet der SETI-Client auf Windows-Systemen als Bildschirmschoner und auf UNIX-Systemen als niedrig priorisierter Hintergrundprozeß. In beiden Fällen sieht man, daß der SETI-Client nur arbeitet, wenn der Computer zur Zeit nicht genutzt wird. Tabelle 4.1 gibt uns einen Eindruck vom heutigen Umfang des SETI-Projektes.

### 4.1.2 Übermittlung der Daten

Wie funktioniert nun die Übermittlung der Daten vom Radioteleskop zu den angeschlossenen Benutzern und wieder zurück? Sehen wir uns dazu das Schaubild 4.1 an (aus [6]).

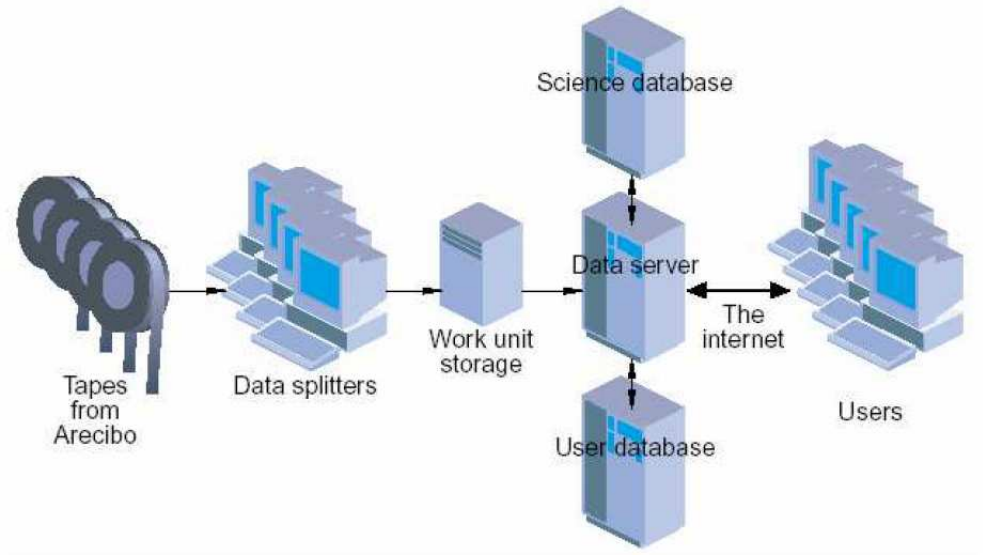
Die vom Radioteleskop auf Band aufegezeichneten Daten werden von Splittern in kleine Datenpakete zerlegt, welche jeweils rund 107 Sekunden aufgezeichnete Daten sowie zusätzliche Informationen über den Ursprung des Signals enthalten. Ein einzelnes Datenpaket (engl. *work unit*) hat die Länge von circa 340 kB.

Diese Datenpakete werden an einen der drei Datenbankserver weitergereicht. Dabei haben die Datenbankserver folgende Funktion:

**User database:** Dieser Server verwaltet die Informationen über die rund 4,5 Millionen Benutzer sowie Statistiken

**Data server:** Er verwaltet die aufgezeichneten Daten sowie mehrere Milliarden (möglicherweise) entdeckte Signale

Abbildung 4.1: Übermittlung der Daten beim SETI@home-Projekt




**Science database:** Diese Datenbank dient als temporärer Speicher für die von Nutzern bearbeiteten work units, sowie als Empfänger von Ergebnissen und Herausgeber neuer work units.

Der Austausch der Daten zwischen Datenbank-Server und Nutzer erfolgt hier über das HTTP-Protokoll, weshalb man das SETI@home-Projekt auch zum "*Internet Computing*" zählt.

Internet Computing

### 4.1.3 Sicherheit und Zuverlässigkeit

Das SETI@home-Projekt sieht mehrere Mechanismen vor, die der Sicherheit und Zuverlässigkeit der ausgewerteten Daten dienen sollen. So werden einzelne work units beispielsweise zeitversetzt an mehrere Nutzer gesendet. Dies soll verhindern, daß Nutzer falsche oder nur unvollständig bearbeitete work units zurücksenden.[2] Ein weiterer Sicherheitsmechanismus hilft zu vermeiden, daß Signale übersehen werden. Zu diesem Zweck werden die work units so angelegt, daß sie sich jeweils um einige Sekunden überlappen.

[2]Die Motivation für ein solches (Fehl-)Verhalten der Nutzer könnte sein, ihre persönliche Statistik, die im Internet von jedermann abgerufen werden kann und Aufschluß über den Beitrag eines Nutzers zum SETI@home-Projekt gibt, zu beschönigen.

Nutzer sind nicht gezwungen, die erhaltenen work units auch tatsächlich vollständig berechnen zu lassen. Deshalb werden work units mehrfach verschickt und erst nach dem Empfang mehrerer Resultate als erledigt gekennzeichnet. Auf diese Weise ist das Projekt unempfindlich gegenüber dem Ausfall einzelner Rechner[3]. Sollte die Verbindung zum Server nicht aufgebaut werden können, so versuchen die Clients jede Stunde erneut, die Verbindung herzustellen, und sind in der Zwischenzeit inaktiv. An dieser Stelle sei erwähnt, daß der Aufbau der Verbindung zum Server stets vom Client initiiert wird, niemals aber vom Server zum Client. Dies ist ein weiterer Beleg dafür, daß einzig der Teilnehmer bestimmt, wann sein Rechner etwas für das Projekt beitragen soll.

Nicht unerwähnt bleiben sollte in diesem Abschnitt auch die Fehlertoleranz des Hauptservers. Zum einen speichert er die Daten redundant und sichert sie regelmäßig. Zum anderen vermerkt er, welcher User welche work unit bekommen hat und verhindert auf diese Weise, daß jemand (gefälschte) Resultate einer work unit zurücksendet, die er gar nicht erhalten hatte. Mit diesem Mechanismus können gültige von ungültigen Daten mit nahezu hundertprozentiger Sicherheit unterschieden werden.

### 4.1.4 Bisheriger Erfolg

Leider(?) hatten bislang alle auffälligen Daten bei näherer Überprüfung ihren Ursprung doch auf der Erde. Aber selbst, wenn sich einmal ein Signal finden ließe, welches nicht irdischen Ursprungs ist, so wäre allein aufgrund der Auswertung der Radioteleskop-Daten von Arecibo noch lange kein Beweis für die Existenz von Außerirdischen gegeben.

Vielmehr würden in diesem Fall Observatorien weltweit so geschaltet werden, daß sie das auffällige Signal ebenfalls empfangen. Man würde sich um eine genauere und langfristigere Aufzeichnung des Signals bemühen, und ein sogenanntes *International Astronomical Union Telegram* (IAUT) veröffentlichen, welches Wissenschaftlern und Privatpersonen zugleich zugänglich ist. Verschiedene Organistationen müßten das Signal als bedeutsam bestätigen.

Eine besondere Vereinbarung, die "Declaration of Principles concerning activities following the detection of Extraterrestrial Intelligence" definiert übrigens den anmessenen Umgang mit der Entdeckung außerirdischer Signale. So sieht sie zum Beispiel vor, daß keine Antwort gesendet werden darf, bevor

[3]Bei rund 4,5 Millionen teilnehmenden Rechnern ist der Ausfall einiger von ihnen wohl alltäglich...

dies von internationalen Gremien befürwortet wird. Der genaue Inhalt dieser Deklaration kann in [7] nachgelesen werden.

### 4.1.5 Zu guter Letzt...

Als die Wiedereröffnung des Arecibo Radio Observatory nach einer Modernisierung im November 1976 gefeiert wurde, wurde eine komprimierte Nachricht gesendet, welche die Repräsentation einer Doppel-Helix, eines Diagramms unseres Sonnensystems sowie einfache Bilder eines Menschen und des Arecibo-Teleskops enthielt. Man wollte auf diese Weise außerirdische Spezies auf uns aufmerksam machen und ihnen die Möglichkeit bieten, mit uns Kontakt aufzunehmen. Bis heute hat sich aber noch niemand gemeldet...

## 4.2 IBM ZetaGRID

Auch das Projekt ZetaGRID, das von IBM ins Leben gerufen wurde, möchte die brachliegende Rechenleistung von ungenutzten Arbeitsplatzrechnern verwenden.

Die Funktionsweise des ZetaGRID ist im Wesentlichen die selbe, wie bei SETI@home: das Client-Programm läuft als Bildschirmschoner bzw. niedrig priorisierter Prozeß auf Windows- oder UNIX-Systemen, und es lädt sich Aufgaben zur lokalen Berechnung herunter. Auch dieses Projekt sorgt für volle Sicherheit sowohl bei der Kommunikation (durch Verschlüsselungsverfahren)[4] als auch bei der Übermittlung der Aufgaben und Ergebnisse (durch Authorisierungsverfahren). Zur Zeit nehmen etwa 500 Rechner am ZetaGRID teil.

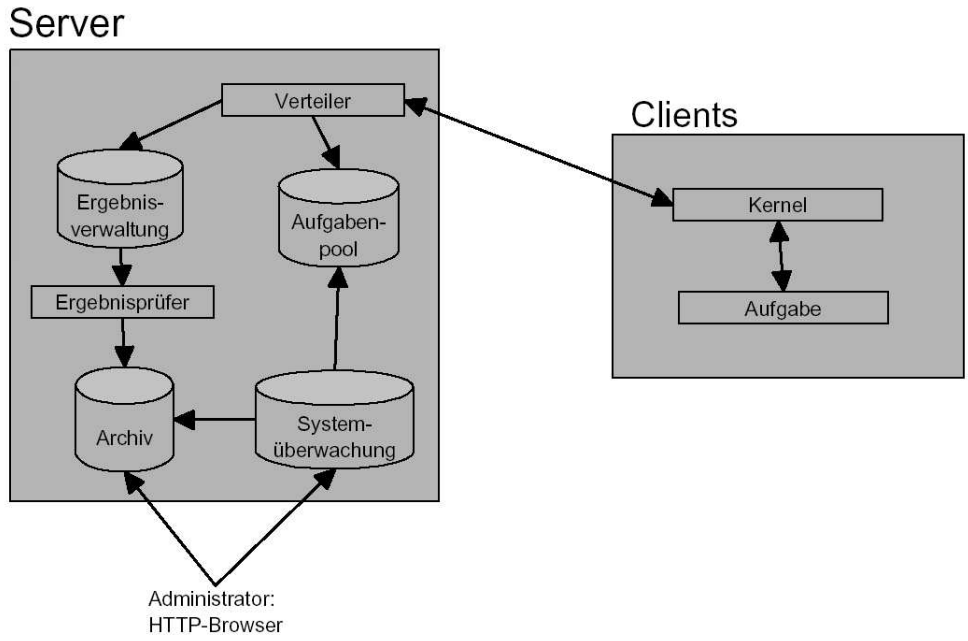
Die Aufgabe, die dem ZetaGRID-Projekt seinen Namen gegeben hat, ist es, Nullstellen der Riemannschen Zeta-Funktion

$$\zeta(s) := \sum_{k=1}^{\infty} \frac{1}{k^s}, \quad s \in \mathbb{C},\ \mathrm{Re}(s) > 1$$

zu finden. Die Riemannsche Hyptohese ist eines der bedeutendsten Probleme in der modernen Mathematik. Sie besagt, daß alle bis auf die triviale Nullstellen der Zeta-Funktion auf der kritischen Geraden liegen, also den Realteil $\frac{1}{2}$ besitzen. Die Nullstellenberechnung der Zetafunktion hat viele Anwendungen in der Zahlentheorie, genauer gesagt in der Kryptographie.

[4] Dies ist nach meinem Kenntnisstand beim SETI@home-Projekt derzeit nicht der Fall

Abbildung 4.2: Komponentenarchitektur ZetaGRID (aus: [9])




Das Clay Mathematics Institute hat die Summe von $ 1.000.000 für den Beweis der Riemannschen Hypothese ausgesetzt, nachdem es in den letzten 150 Jahren nicht gelungen ist, diese Vermutung entweder zu beweisen oder zu widerlegen. Die Komponentenarchtitektur, die wir in Abbildung 4.2 sehen, weist eine starke Ähnlichkeit zum SETI@home-Projekt auf:

An dieser Stelle sei jedoch erwähnt, daß die Berechnung der Nullstellen lediglich ein Anwendungsbeispiel für das ZetaGRID darstellt. Im Unterschied zu SETI@home dient ZetaGRID also nicht bloß der Berechnung einer komplexen Aufgabe. Vielmehr bietet IBM die im ZetaGRID verwendete Technik auf dem Markt als Dienstleistung für Projekte an, die sich zu ihrer Lösung des Grid-Computings bedienen möchten.

## 4.3 CERN-Data-Grid (LHC)

Das dritte Beispiel in diesem Kapitel führt uns zu einem Data-Grid[5]. Wir werden uns im Folgenden nun das *LHC-Computing-Grid-Projekt* etwas näher ansehen. LHC steht dabei für *Large Hadron Collider*, worunter ein derzeit im Bau befindlicher Teilchenbeschleuniger, der einmal der weltweit leistungsfähigste seiner Art sein soll, zu verstehen ist. Seine Fertigstellung ist für das Jahr 2005 geplant. Mit Hilfe des LHC soll der Ursprung der Masse untersucht und damit eine Antwort auf eine der grundlegendsten Fragen der Wissenschaft gegeben werden. Der LHC entsteht derzeit am CERN, der europäischen Organisation für nukleare Forschung, mit Sitz in Genf (Schweiz).

Das CERN-Institut erwartet aus dem Einsatz des LHC ein immenses (und bisher noch nicht gekanntes) Datenaufkommen[6], zu dessen Bewältigung es Computational Grids einsetzen möchte. Genauer "sollen weltweit freie oder ungenutzte Prozessor- und Speicherkapazitäten von Zehntausenden von Computern zu einem Datennetzwerk (sog. DataGrid) zusammengefaßt werden"[3]. Dabei ist die Idee, die Daten nicht wie bisher an einem Ort vorzuhalten und zu verarbeiten, sondern auf mehrere Zentren zu verteilen uns transparent zu nutzen[10]. Dazu soll ein in Abbildung 4.3 dargestelltes hierarchisches System mit verschiedenen Ebenen (*Tiers*) verwendet werden. In diesem Modell stellt das CERN die oberste Ebene (Tier-0) dar. Seine Aufgabe besteht darin, Experimentdaten zu speichern und zu rekonstruieren. Die Analyse und die Erzeugung simulierter Ereignisse wird dann in regionalen Tier-1-Zentren vorgenommen, an die sich in einer weiteren Hierarchie-Ebene (nationale) Tier-2-Zentren anschließen. Die vierte und fünfte Ebene bilden sodann Institutsrechner (Tier-3) sowie einzelne Arbeitsplatzrechner (Tier-4).

Der Planung nach sollen die Regionalzentren in Italien, Frankreich, Großbritannien, Japan und in den USA entstehen. Mit Hilfe des Data-Grids (und letztlich auch eines Computational Grids) kann erreicht werden, daß alle beteiligten Wissenschaftler in den Regionalzentren arbeiten können und dies nicht eigens am CERN selbst machen brauchen.

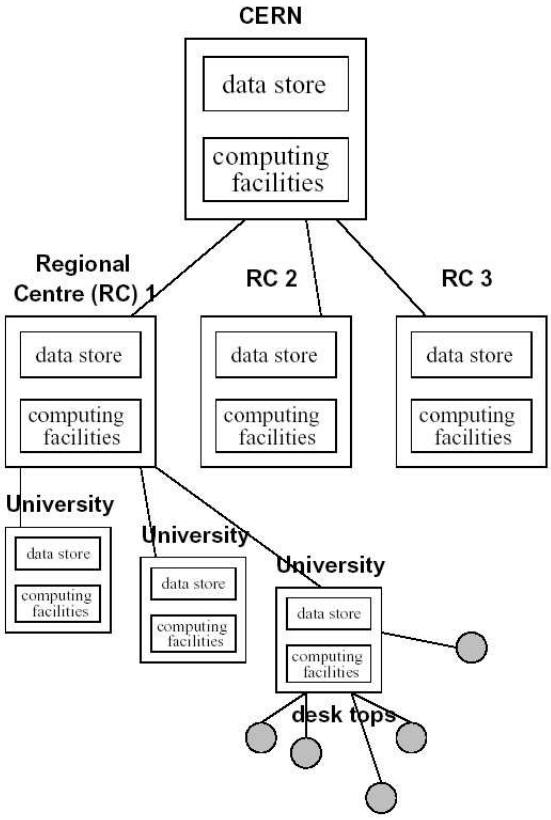
## 4.4 Weitere Projekte

Die vorstehend vorgestellten Grid-Computing-Projekte sind nur drei Beispiele. Die Anzahl derartiger Projekte wächst zunehmend an. Ein weiteres solches

[5] Auch wenn hier ein Computational Grid und ein Data Grid zusammfallen, so wollen wir das Projekt aus dem Blickwindel des Data-Grids betrachten.

[6] Mehrere Petabytes Daten jährlich über eine Lebenszeit von 15 bis 20 Jahren[11]

Abbildung 4.3: Geplantes LHC-Netzwerk am CERN (aus [11])



ist die Suche nach Mersenn'schen Primzahlen[7]. Motivation für die Teilnahme an dieser Suche ist unter anderem gewiß auch das Preisgeld von $100.000 für den Entdecker einer Mersenn'schen Primzahl, die bestimmten Anforderungen genügt. Für weitere Informationen zu diesem Thema sei auf die Internet-Seite `http://www.mersenne.org` verwiesen.

Mit den in diesem Kapitel vorgestellten Beispielen soll nur ein Einblick in existierende Grid-Computing-Projekte gegeben werden. Insoweit soll an dieser Stelle darauf verzichtet werden, alle derzeit bekannte bzw. existierende Projekte aufzuzählen.

[7]Dieses Projekt trägt den Namen "*GIMPS*" *(The Great Internet Mersenne Prime Search).*

# Kapitel 5

# Abschließende Bemerkungen

## 5.1 Grid-Computing und seine Beziehungen zu anderen Technologien

Wo könnte in Zukunft Grid-Computing eingesetzt werden. In [12] finden sich unter anderem folgende Vorschläge:

### 5.1.1 World Wide Web

Das World Wide Web ist mitterweile fast überall verfügbar. Seine verwendeten Technologien (TCP/IP, HTTP, HTML, XML, etc.) dienen als ideale Basis, um virtuelle Organisationen aufzubauen. Wenngleich die Interaktion Browser/Client mit WebServer problemlos funktioniert, gehen diese Technologien mitunter für anspruchsvollere Aufgaben, die in virtuellen Organisationen denkbar sind, nicht weit genug. An dieser Stelle könnte Grid-Computing helfen.

### 5.1.2 Application und Storage Service Providers

Application Service Providers (ASP) bieten ihren Kunden Zugriff auf eine bestimmte Hard- und Softwarekombination an, Storage Service Providers (SSP) bieten hingegen Speicherplatz an.

Für diese Form der Dienstleistung kann Grid-Computing sehr interessant werden. Im Falle der ASP kann eine Trennung von Hard- und Software erreicht werden, so daß der Kunde zukünftig seine eigene Software auf Hardware des Providers ausführen kann. Denkbar ist auch die andere Richtung:

der Kunde führt Software, die auf dem Grid zur Verfügung steht, auf seiner eigenen Hardware aus. Ebenso kann er die Dienste der ASP mit jenen der SSP kombinieren.

### 5.1.3 Peer-to-Peer- und Internet-Computing

Wer hat nicht schon einmal mit Peer-to-Peer-Programmen wie z.B. Napster oder Gnutella gearbeitet? Letztlich sind sie eine Vorstufe des Grid-Computings, denn sie ermöglichen unkompliziert den Austausch von Dateien, also das File-Sharing, welches bereits ein Bestandteil des Grid-Computings darstellt. Es handelt sich hier genauer gesagt um ein Beispiel für ein ganz einfaches Daten-Grid.

Ein weiterer bereits in Anwendung befindlicher Teil des Grid-Computings bezieht sich auf das Internet-Computing. Hierunter versteht man das, was im SETI@home-Projekt, welches wir in Abschnitt 4.1 auf Seite 29 bereits vorgestellt haben, zur Anwendung kommt.

## 5.2 Aussagen zum Grid-Computing und deren Bewertung

Das Grid Computing befindet sich derzeit noch in der Entwicklung. Deshalb gibt es viele Meinungen und Aussagen über das Grid. Nachstehend zitieren wir drei Statements und deren Kommentierung (sinngemäß, gekürzt) durch Ian Foster, einem der Pioniere des Grid-Computings (vgl. [12], Seite 17ff).

**Aussage 1:** Das Grid ist das Internet der nächsten Generation!

**Kommentar:** Das Grid wird keine Alternative zum Internet werden. Vielmehr ist es eine Erweiterung des Internets durch mehrere neue Protokolle und Dienste, die auf den Internet Protokollen aufsetzen.

**Aussage 2:** Das Grid erfordert ein verteiltes Betriebssystem!

**Kommentar:** Grid Software sollte die (zusätzlichen) Dienste der Betriebssysteme definieren, die zukünftig auf allen Rechnern vorhanden sein sollen. Diese Dienste leisten für das Grid dann das, was das Betriebssystem für den einzelnen Computer leistet.

**Aussage 3:** Das Grid wird High-Perfomance-Computer überflüssig machen!

**Kommentar:** Die hundert, tausend oder sogar Millionen von Prozessoren, die innerhalb einer virtuellen Organisation verfügbar sein werden, repräsentieren eine signifikante Quelle der Rechenpower. Deshalb werden High-Performance-Computer aber nicht unbedingt überflüssig. Viele Probleme erfordern eng zusammenarbeitende Computer mit geringen Zugriffszeiten und großer Kommunikations-Bandbreite. Deshalb wird Grid-Computing die Nachfrage für Großrechner eher erhöhen als senken, weil es den Zugriff auf solche Großrechner erleichtert.

## 5.3 Probleme des Grid-Computings

Selbstverständlich hat auch das Grid-Computing seine Schattenseiten. Insbesondere bei der Realisierung der Implementierungen sieht man sich schnell vor Probleme gestellt. Das Global Grid Forum[14] nennt an dieser Stelle unter anderem folgende Probleme:

- Es existiert eine sehr große Vielfalt von Systemen.
- Oft sind extrem hohe Übertragungskapazitäten notwendig.
- Das Vertrauen der Rechenzentren und Dienstleistungsanbieter muß garantiert sein
- Es gibt derzeit noch viele offene Probleme bei Standards und Programmierung

# Index

Agenten, 8
Anwendung, 9
application, 9
authentication, 27

CERN, 35
checkpointing, 12
Cluster, 19
Clustern, 9
Connectivitylayer, 23

Datastriping, 13
Delegation, 27
donor, 25
Dynamik, 8

Fabriclayer, 22

Grid, 4, 7
GUI, 26

IAUT, 32
IBM, 33
Intergrid, 20
interoperability, 22
Intragrid, 20

Jobs, 9
jobscheduler, 16

Konsistenz, 18

LHC, 35

middleware, 22
MPI, 21

Policies, 8

replication, 13
reservation, 15

scavenging, 15
scheduling, 15
SETI, 29
Skalierbarkeit, 11
Subjobs, 9

Tiers, 35

Zetafunktion, 33
ZetaGRID, 33

# Abbildungsverzeichnis

2.1 Anwendungen, Jobs und Subjobs (aus [4]) . . . . . 10
2.2 Data Striping (aus [4]) . . . . . 14

3.1 Gegenüberstellung Grid- bzw. Internet Protocol Architecture (aus [12]) . . . . . 23

4.1 Übermittlung der Daten beim SETI@home-Projekt . . . . . 31
4.2 Komponentenarchitektur ZetaGRID (aus: [9]) . . . . . 34
4.3 Geplantes LHC-Netzwerk am CERN (aus [11]) . . . . . 36

# Tabellenverzeichnis

4.1 Nutzerstatistik bei SETI@home, Stand: 20.06.2003, aus [8] . . 30

# Literaturverzeichnis


[1] Fraunhofer Magazin 1.2003, Seite 14 f

[2] Grid Computing Info Centre, `http://www.gridcomputing.com`

[3] Nachrichten - Forschungszentrum Karlsruhe, Jahrgang 34, 4/2003, S. 280-283

[4] IBM - Redbook "Fundamentals of Grid Computing", Viktors Berstis, IBM 2002, `http://www.ibm.com/redbooks` (letzter Zugriff 26.07.2003)

[5] "SETI, Sharing Terrestrial Intelligence", Daniel Görlich, Humboldt-Universität Berlin, Mai 2002

[6] Schaubild Datenübertragung, aus: `http://www.computer.org/cise/articles/seti.htm` (letzter Zugriff 21.06.2003)

[7] "Declaration of Principles concerning activities following the detection of Extraterrestrial Intelligence", `http://www.seti.org/science/principles.html` (letzter Zugriff 21.06.2003)

[8] Homepage des SETI@home-Projekts `http://setiathome.ssl.berkeley.edu/totals`

[9] Grid Computing, Die IBM ZetaGRID Lösung, Ralf Grohmann und Dr. Sebastian Wedeniwski, `http://www.zetagrid.net/zeta/ZetaGRID4costumers_de.pdf` (letzter Zugriff 26.07.2003)

[10] Report of the Steering Group of the LHC Computing Review, CERN/LHCC/2001-004, `http://lhc-computing-review-public.web.cern.ch/lhc-computing-review-public/Public/Report_final.pdf`

[11] Data Management in an International Grid Project, Wolfgang Hoschek et al.
`http://citeseer.nj.nec.com/hoschek00data.html` (letzter Zugriff 26.07.2003)

[12] The Anatomy of the Grid, Ian Foster et al., `http://www.globus.org/research/papers.html#anatomy`

[13] Design anf Deloyment of a National-Scale Authentication Infrastructure, Butler, R. et al., IEEE Computer, 33(12):60-66, 2000

[14] Global Grid Forum (GGF), `http://www.gridforum.org`

# Semantic Web

Seminar „Internet Dienste“
SS 03

von
Stefanie Beiler

# 1. Problemstellung

Das World Wide Web hat sich in den letzten Jahren zu einer Informationsquelle mit unbegrenzten Potential entwickelt. Doch dieser ‚größten Datenspeicher der Welt', den täglich Millionen von Usern nutzen, hat ein Problem: nicht zu bändigende Unordnung. Suchmaschinen versuchen zwar, Ordnung in dieses Chaos zu bringen, doch sie stoßen immer wieder an die Grenzen mit der Frage: welche Seite liefert nun das, was der User sucht?

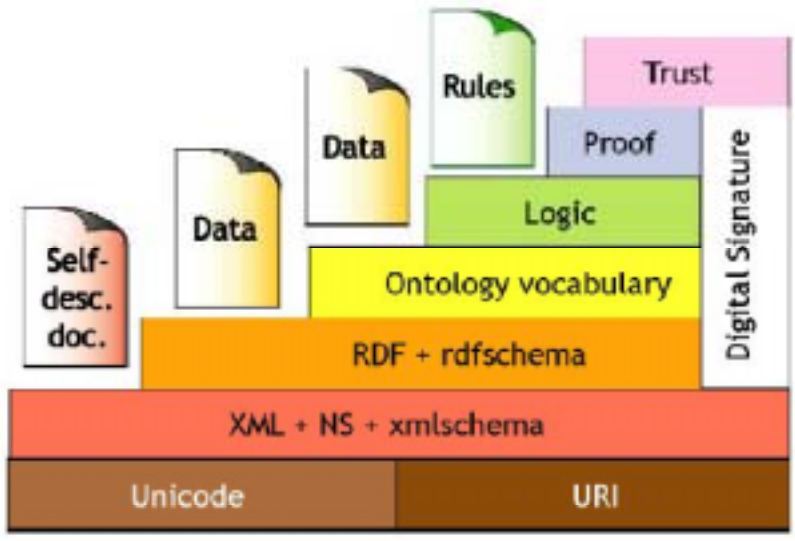
Ein Problem der Suchmaschinen ist zugleich der Vorteil des Web: seine Unbegrenztheit. „Google" gab bei der Suche nach „Semantic Web" über 230.000 Treffer an, von denen viele keine relevanten Informationen liefern. Das Problem dahinter ist, dass Internet-Seiten genaue Informationen darüber enthalten, wie ihre Inhalte dargestellt werden (Schriftart, Schriftgröße, Hintergrundfarbe usw.), aber keine genaueren Daten über ihren Inhalt. Die Suche basiert daher nur auf Wort-Vektoren und gestaltet sich wie die Suche nach der Nadel im Heuhaufen.

Um Daten aber maschinell sinnvoll verarbeiten zu können, d.h. dass sie sowohl von Menschen als auch von Maschinen verstanden werden können, bräuchte man Informationen, in welchem Zusammenhang die Daten stehen, also Daten über Daten, sogenannte Meta-Daten.

Das World Wide Web Consortium (W3C) und einer seiner Gründer, Tim Berners-Lee, haben dieses Problem schon vor langer Zeit erkannt, und ein Modell entwickelt, das Informationen im Internet ihrer Bedeutung nach klassifiziert und einander zugeordnet. Diese Lösungsansätze wurden unter **Semantic Web** zusammengefasst.

# 2. Grundidee des Semantic Webs

Die Idee des Semantic Webs ist, dass Maschinen selbstständig Daten im Netz verwerten, um nutzbringende Informationen herauszufiltern. Seine Vorstellung, wie diese Idee verwirklicht wird, stellt Berners-Lee in einem „layer-cake" dar:




© Tim Berners-Lee
Semantic Web on XML,

6.12.2000, Washington DC

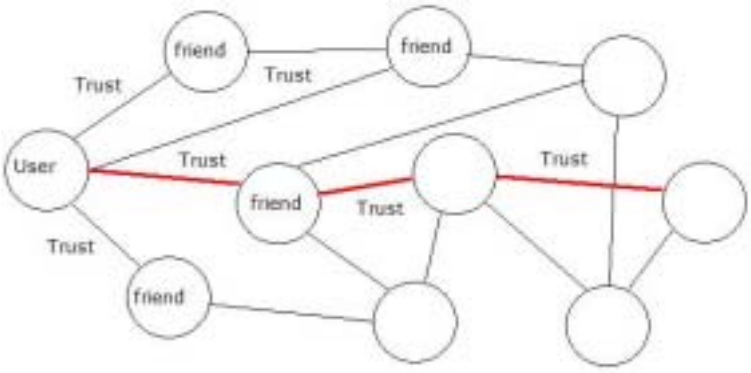
Das Semantic Web wird auf das existierende Web aufgebaut und wird somit mit anderen Webtechnologien interagieren. Der Inhalt jeder Webseite steht im Unicode (ISO 10646), ein universelles Zeichenrepertoire, dass alle Zeichen der Welt umfasst und ständig erweitert wird. Jede Ressource im Web hat eine URI (Uniform Resource Identifier), die den eindeutigen Namen und den eindeutigen Pfad dieser Ressource darstellt. Eine bekannte Form der URI ist die URL.
XML, Namespace, RDF und Ontologien beschreiben den Inhalt näher (siehe dazu Kapitel 3).

Die nächsten Stufen dieses „layer-cakes“ befinden sich noch in der Entwicklung. Wichtig ist es, eine Logik in der Semantik zu bringen. Ein aktuelles Beispiel dazu wären Rabattmarken. Ein Kunde erhält, nachdem er für 1000 € eingekauft hat, 5% Rabatt. Die Maschinen muss dann erkennen, dass, wenn er für einmal für 527€ und einmal für 498€ eingekauft hat, er auch 5% bekommt. Die menschliche Logik ist dabei natürlich wichtiger als die rationale.

In der Zukunft spielt es natürlich auch eine wichtige Rolle, Beweise für den Inhalt dieser Informationen zu haben. Die Grundfrage ist hierbei: Wenn eine Quelle behauptet, X ist blau, und eine andere sagt, X ist nicht blau, fällt dann das Semantic Web auseinander? Antwort: NEIN!!

1. Aussagen beruhen i.A. auf den Zusammenhang
2. Aussagen sollen in der Zukunft durch Beweis-Algorithmen und die digitale Signatur überprüft werden

Diese Beweis-Algorithmen sind schwierig zu entwickeln, da viele Verbindungen überprüft werden müssen.




Berner-Lee’s Wunschvorstellung ist ein „Web of Trust“. Jeder User hat Freunde und Bekannte, die er kennt und auf deren Meinung er vertaut, z.B. über einen Film. Diese Freunde haben wiederum Freunde, denen sie vertrauen, usw. Somit vertraut man also auch allen Bekannten des eigenen Freundes.
Denkbar ist es, verschiedene Vetrauensstufen einzuführen, die die Länge der Vertrauenskette wiederspiegeln.

# 3. Sprachen des Semantic Webs

## 3.1. XML (Extensible Markup Language)

XML wurde entworfen, um auf einfache Weise Dokumente im Web zu verschicken. Jeder kann eigenes Dokumentenformat erstellen und danach ein Dokument in diesem Format erstellen. Dieses Format kann Markierungen enthalten, die den Inhalt des Dokuments beschreiben. Diese Markierungen können dann von Maschinen gelesen und verarbeitet werden.
Ein Beispiel für ein solches Dokument ist:

```
<sentence>
     <person>Ich</person> habe eine <animal>Katze</animal>.
</sentence>
```

Man erkennt, dass XML eine familiäre Syntax zu HTML hat, sogenannte Tags. Es gibt einen öffnenden (`<sentence>`) und einen schließenden (`</sentence>`) Tag, die dem Computer mitteilen, dass `Ich habe eine Katze` ein Satz ist. Während man HTML mehr für die visuelle Präsentation benutzt, verwendet man XML, um Daten und ihre Bedeutung zu beschreiben.

Manchmal ist es sinnvoll, dem Computer mehr Informationen zu geben, z.B. welche Person gemeint ist, indem man Attribute hinzufügt. Das ganze sieht dann folgendermaßen aus:

```
<sentence>
     <person href="http://aaron.com">Ich</person> habe eine
     <animal>Katze</animal>.
</sentence>
```

Um sicherzustellen, dass XML-Dokumente eine korrekte Struktur besitzen, gibt es Schemata. Die Document Type Description DTD definieren eine Grammatik für XML-Dokumente und schränken eine beliebige Verschachtelung ein. Zu DTD's abwärtskompatible sind XML Schemas mit komplexe Datendefinitionssprachen und Typisierungen.

## 3.2. XML Namespaces

Da XML Daten strukturiert, wird es vor allem Programmen für den Datenaustausch übers Internet benutzt. Doch was passiert, wenn ein Programm versucht, zwei XML-Dokumente zu verschmelzen, die gleiche Tags mit unterschiedlichen Bedeutungen benutzen? `<sentence>` kann in einem englischen Text nicht nur für „Satz" stehen, sondern auch „Gefängnisstrafe" bedeuten.
Um solchen Konflikten vorzubeugen, benutzt man sogenannte XML Namespaces. Dabei wird einem Element bzw. Tag eine URI zugeteilt, die dann das Element beschreibt und eindeutig identifiziert. Durch Verwendung von Präfixen ist die gewünschte Eindeutigkeit der Tags wiederhergestellt.
Falls z.B. <news> und </news> doppelt vorkommt, hilft das Voranstellen eines eindeutigen Präfix:

`<cnn:news>...</cnn:news>` und `<n24:news>...</n24:news>`

Teilt man dann noch jeweils eine URI bzw. URL zu, erhält man

```
<cnn:news  xmlns:cnn="http://www.cnn.com/xmlns/">…</cnn:news>
<n24:news  xmlns:n24="http://www.n24.de/">…</n24:news>
```

Weitere auf XML basierende Technologien, die XML für viele Anwendungsbereiche interessant macht:

- XML Stylesheets
- Cascading Stylesheets
- XSL
- Xpath
- XQL/XML-QL

Des weiteren ist noch zu erwähnen, dass XML als W3C-Technologie lizenzfrei ist, plattformunabhängig und in der Industrie unterstützt wird.

## 3.3. RDF (Resource Description Framework)

Damit Maschinen lesbare und eindeutig verarbeitbare Aussagen über Ressourcen machen können, ist XML nicht mehr ausreichend. Ziel von RDF ist es, dass Metadaten über eine Ressource in einer Form dargestellt werden, bei der der Sinn der Aussage bei der Übertragung im Internet nicht verloren geht.

Jedes RDF Statement (eine Aussage in RDF) hat Ähnlichkeit mit einem einfachen Satz, nur das fast alle Wörter URIs sind. Jedes Statement hat drei Teile: *subject*, *predicate* und *object*. Um zum Beispiel auszudrücken, dass Aaron das Buch „Weaving the Web“ von Tim Berners-Lee mag, erhält man das RDF Statement:

```
<http://aaron.com/>
<http://love.example.org/terms/reallyLikes>
<http.//www.w3.org/People/Berners-Lee/Weaving/>
```

Bei diesem Beispiel wurde eine sehr einfache Notation verwendet, die sogenannte **N-Triples Notation**. RDF kennt noch andere Notationen, u.a. die Notation als gerichteter Graph oder die von W3C offizielle empfohlene Notation in XML (RDF/XML), die etwas komplizierter ist, aber im Grunde das gleiche aussagt:

```
<rdf:RDF
     xmlns:rdf=“http://www.w3.org/1999/02/22-rdf-syntax-ns#“
     xmlns:love=“http://love.example.org/terms“
>
     <rdf:Description rdf:abaout=“http://aaron.com/">
          <love :reallyLikes
rdf :resource="http://www.w3.org/People/Berners-Lee/Weaving/” />
     </rdf:Description>

</rdf:RDF>
```

Man erkennt nun folgende Vorteile von RDF:

- Anpassbarkeit & Flexibilität: beliebige Objekte lassen sich beschreiben. Auf welche Eigenschaften man dabei zurückgreift, bleibt einem selbst überlassen.
- Austauschbarkeit: da empfohlen wird, RDF in XML zu notieren, ist es ziemlich einfach, RDF-Dokumente auszutauschen. Die Bedeutung des Inhalts wird direkt in die Webpage mit eingebunden, ohne vom Browser angezeigt zu werden.
- Einfachheit: RDF basiert auf Statements, die aus drei Komponenten bestehen. Dadurch ist es ziemlich einfach, RDF weiterzuverarbeiten
- Kombinierbarkeit: subject, predicate und object können alle selbst wieder Ressourcen sein. Das erlaubt, sie zu beschreiben und miteinander zu verschachteln.

RDF ist also bestens geeignet, Aussagen über Dinge zu machen. Die Quelle dieser Aussagen muss aber keineswegs der Mensch sein. Intelligente Programme können Datenbanken im Internet verarbeiten, Daten zusammenfügen und dann Schlüsse daraus ziehen.

### 3.3.1. Bestehende Anwendungen von RDF

**Dublin Core**[1]
Die bekannteste Anwendung von RDF ist das sogenannte Dublin Core Element Set. Es besteht aus 15 Metadatenelementen zur inhaltlichen und formalen Beschreibung von Internetquellen. Ziel ist eine verbesserte Indizierung und dadurch auch ein verbessertes Retrieval. Forschungseinrichtungen des Bibliothekwesens erstellten das Dublin Core Schema, das entsprechend den jeweiligen Bedürfnissen des Anwenders erweiterbar ist.

**Fotoalbum**[2]
2000 beschrieben Yves Lafon und Bert Bos, Mitarbeiter von W3C, ein Projekt, dass sich mit Beschreibung und Wiederauffindung von digitalisierten Fotos beschäftigt. Die Metadaten werden dabei in die Bilddatei in einem Kommentarblock des jpg-Dateiformats gespeichert und können auch ohne Bilddaten mittels http angefordert werden. Der Artikel (siehe Fußnote) enthält auch das von Dublin Core abgeleitete Schema für die Beschreibung der Fotos.

**UK-Mirror-Service**[3]
Der Spiegeldienst des Vereinten Königreichs wird gemeinsam von den Universitäten von Kent und von Lancaster betrieben. Dabei handelt es sich um eine Sammlung von mehr als 2 Millionen Programmdateien, die von anderen Orten vollständig kopiert (gespiegelt) wurden. Um diese Dateien zu verwalten und leichter zu durchsuchen wird RDF eingesetzt.

**PICS (Platform for Internet Content Selection)**[4]
PICS ist ein Standard zur Inhaltsbewertung einer Internet-Ressource, wodurch z.B. ungeeignete Inhalte für Kinder zur Filtersoftware gesperrt werden können. Die Bewertung kann durch den Inhaltsanbieter oder andere Organisationen erfolgen. Seit März 2000 können diese Bewertungen mittels RDF erfolgen, was genauere Beschreibungen und Kategorisierungen ermöglicht.

## 3.4. Schemas und Ontologien

Ein Problem besteht jetzt immer noch: Mit RDF und XML sind jetzt zwar durch URIs eindeutige Konzepte möglich und es können auch Aussagen zu diesen Konzepten getroffen werden, doch der Zusammenhang von Konzepten untereinander ist immer noch ungeklärt. Eine Metasprache alleine bringt keine Semantik ins Web. Dazu müssen Schemas und Ontologien definiert werden. Die Spezifizierung von Ontologien ist eines der wichtigsten Aspekte des Semantic Webs.

Schemas und Ontologien bieten folgende Vorteile:

- Bereitstellung einer gemeinsamen Grundlage zum Austausch von Informationen

[1] „http://www.dublincore.org“
[2] “Describing and retrieving photos using RDF and HTTP”, Yves Lafon und Bert Bos, 2000
[3] “http://www.mirror.ac.uk/help/rdf.html”
[4] “PICS Rating Vocabularies in XML/RDF”, W3C Note, Edited by Dan Brickley, W3C and Ralph Swick, W3C, 27. März 2000

Ein Agent kann bei einem Onlineshop genau feststellen, mit welchen Ressourcen er es zu tun hat und ob diese für seine Aufgabe nützlich sind.
- Einfache Wiederverwendbarkeit
- Informationen können von verschiedenen Ressourcen gesammelt und weiterverarbeitet werden.
- Die Trennung von Informationen und deren Verarbeitung ermöglicht es, auf Änderungen der Information besser und einfacher zu reagieren.

Ontologien werden schon in der Entwicklung auf die spätere Benutzung ausgelegt. Deshalb gibt es im Semantic Web nicht nur eine, sondern sehr viele Ontologien, die kleine Anwendungsbereiche haben. Die Entwicklung umfasst dann folgende Bereiche:
- Definition von Klassen in der Ontologie
- Unterbringung von Klassen in einer Ober-/Unterhierarchie
- Definition von Attributen und Beschreibungen der Eigenschaften dieser Attribute
- Erstellung von Instanzen

Dabei gelten folgende Grundsätze:
1. Es gibt keine richtige Lösung, sondern es existieren immer mehrere. Die Wahl der Lösung hängt sowohl von der Anwendung als auch von den möglichen Erweiterungen ab, für die die Ontologie erstellt wird.
2. Die Entwicklung einer Ontologie ist iterativer Prozeß. Im Laufe des Lebenszyklus einer Ontologie wird diese mehrmals abgeändert.
3. Die Konzepte innerhalb einer Ontologie sollen möglichst nahe an der Wirklichkeit angelehnt sein.

RDF ist eine sehr beschränkte Sprache. Erweiterungen wie RDF Schema und DAML+OIL (DARPA Agent Markup Language with Ontology Inference Layer) erweitern RDF und bieten die Möglichkeit, Schemas und Ontologien zu erstellen.

### 3.4.1. RDF Schema

RDF-Schema ist kein Schema im eigentlichen Sinne, sondern erweitert RDF um Konstrukte zur Kreierung von Schemas. Das RDF Schema definiert eine Reihe von Eigenschaften und Klassen, die zur Erstellung von Schemas benutzt werden können. Einige wichtige RDF-Schema Elemente sind nachfolgend aufgeführt:

Basisklassen:
- `rdfs:Resource`: Alle Dinge, die mittels RDF beschrieben werden können, werden als Ressourcen bezeichnet
- `rdf:Property`: Eine Unterklasse von `rdfs:Resource` – das sind jene Elemente, die zum Beschreiben von Ressourcen verwendet werden können
- `rdfs:Class`: Ressourcen können zur besseren Katalogisierung in Klassen eingeteilt werden.
  Werden in einem Schema neue Klassen definiert, so sind diese Unterklassen von `rdfs:Class`

Basiseigenschaften:
- `rdf:type`: Gibt an, dass eine Ressource Mitglied einer Klasse ist. Mögliche Wertebereiche von `rdf:type` sind Instanzen von `rdfs:Class`

- `rdf:subClassOf`: Diese Eigenschaft spezifiziert transitive Beziehungen zwischen Klassen. Mögliche Werte für `rdfs:subClassOf` sind Instanzen von `rdfs:Class`
- `rdfs:subPropertyOf` ist eine Instanz von `rdf:Property`, welche angibt, dass eine Eigenschaft eine nähere Beschreibung einer anderen ist.
- `rdfs:seeAlso`: Gibt eine Ressource an, welche zusätzliche Informationen über die zu beschreibende Ressource enthält.
- `rdfs:isDefinedBy`: Ist eine Untereigenschaft von `rdfs:seeAlso` und gibt jene Ressource an, welche das Subjekt definiert.
- `rdfs:Comment`: Fügt einer Ressource eine Beschreibung zu
- `rdfs:label`: Stellt ein Label für einen Ressourcennamen zur Verfügung

Seine eigenen Klassen zu erstellen ist damit ziemlich einfach, z.B. eine Klasse „Dog“, die alle Hunde dieser Welt enthält:

```
:Dog rdf:type rdfs:Class .
```

rdf:type zeigt, daß Dog Mitglied dieser Klasse ist. Wenn wir jetzt sagen, „Fido“ ist ein Hund, dann ergibt das

```
:Fido rdf:type :Dog .
```

### 3.4.2. DAMl+OIL

DAML+OIL ist eine Ontologiesprache, die die Äquivalenz zweier verschiedener Ressourcen erklären kann. Sie bietet die Möglichkeit, Ontologien zu erstellen und Objekte einer Ontologie mit Objekten einer anderen Ontologie in Beziehung zu stellen. Anwendungen können dadurch mögliche Schlussfolgerungen ziehen und unbekannte Ressourcen in Beziehung mit anderen setzen. Dies ermöglicht dann wiederum die Verarbeitung solch anderer Ressourcen.

Als Beispiel nehmen wir an, dass „Katze“ das Gegenteil von „Hund“ ist. In DAML+OIL ausgedrückt sieht das dann so aus:

```
:dog daml:inverse :cat .
```

# 4. Mögliche Anwendungsbereiche

Das eigentliche Ziel des Semantic Web besteht darin, eine höhere Informationsgehalt zu gewinnen. Bisher von Menschen getätigte Arbeit soll automatisiert und von Maschinen erledigt werden. Dabei kann es sich zum Teil um komplexe und sich immerwährend ändernde Aufgaben handeln.

Es gibt viele Anwendungsbereiche, die vom Semantic Web profitieren oder dadurch erst möglich werden.
Ein solcher Bereich sind Suchmaschinen.

### 4.1. Suchmaschinen

Die aktuellen Suchmaschinen beruhen auf Wort-Vektoren, d.h. es werden alle Dokumente als Treffer ausgegeben, in denen die gesuchten Wörter vorkommen, unabhängig von der Sprache oder Bedeutung.
Beim Semantic Web der Zukunft ist es denkbar, falls die Ergebnismenge zu groß ist, man sie weiter einschränken kann, allerdings nicht, indem man weiter Suchbegriffe hinzufügt, sondern man die ursprünglichen Wörter genauer spezifiziert. „Koch“ soll dann z.B. eine Name sein und kein Beruf.
Die Suche ist dann nicht mehr nur rein lexikalisch, sondern auch semantisch eingeschränkt.

### 4.2. Agenten

Eine anderes Anwendungsgebiet stellen Agenten dar, die sich selbstständig im Internet bewegen und verschiedene Aufgaben erledigen. Die Suche nach einem passenden Arzt oder das Finden von Terminen wird von diesen Agenten selbstständig ausgeführt.
Für solche Agenten ist das Semantic Web nicht die einzige Voraussetzung, aber es vereinfacht die Entwicklung enorm.

# 5. Ausblick & Kritik

Obwohl RDF mittlerweile immer öfter benutzt wird, und XML von Mozilla/Netscape und dem im Herbst erscheinendem neuen Office von Microsoft unterstützt wird, ist das hier beschriebene Semantic Web eine Vision und noch nicht weit verbreitet. In Zukunft wird es aber zweifellos an Bedeutung gewinnen. Das Internet wächst unaufhaltsam und man braucht Anwendungen, die sich im Web selbstständig bewegen können, um seine Ressourcen vollständig auszuschöpfen.

Man darf das Semantic Web aber auch nicht als Allheilmittel ansehen. Es ist keine triviale Aufgabe, Dokumente in RDF zu beschreiben, da dies mehr oder weniger auf das Herausfinden von Schlüsselwörter reduzierbar ist, und das selten zu den gewünschten Resultaten führt. Ein weiterhin unüberbrückbares Problem ist die Unendlichkeit des Internets.

Die Vision eines Agenten, dem man nur noch sagen muss, wo und wie lange man Urlaub machen will, der dann selbstständige die günstigsten Angebote durchsucht, ein Flug und Hotelzimmer bucht, ein Auto mietet und solange die Zeitung abmeldet, liegt noch in weiter Ferne. Aber das Semantic Web ist immerhin ein Grundstein dafür.

# 6. Linksammlung

1. World Wide Web Consortium: www.w3.org/
2. W3C Semantic Web: www.w3.org/2001/sw/
3. The Semantic Web Community Portal: www.SemanticWeb.org/
4. Seminar Grundlagen des Semantic Web: www.inf.fu-berlin.de/inst/ag-nbi/lehre/03/S_SW/
5. Resource Description Framework (RDF) Model and Syntax Specification: www.w3.org/TR/1999/PR-rdf-syntax-199900222

# 7. Inhaltsverzeichnis


1. Problemstellung..... 2
2. Grundidee des Semantic Web..... 2
3. Sprachen des Semantic Web..... 3
   3.1. XML..... 3
   3.2. XML Namespace..... 4
   3.3. RDF..... 5
      3.3.1. Bestehende Anwendungen von RDF..... 6
   3.4. Schemas und Ontologien..... 6
      3.4.1 RDF Schema..... 7
      3.4.2. DAML+OIL..... 8
4. Mögliche Anwendungsbereiche..... 8
   4.1. Suchmaschinen..... 9
   4.2. Agenten..... 9
5. Ausblick & Kritik..... 9
6. Linksammlung..... 10

Universität Ulm

# Seminar Internet Dienste

# Webservices

Matthias Kirchmayr, SS 2003

# Inhaltsverzeichnis

**1 Motivation** **1**

**2 Definition** **1**

**3 XML & Co.** **3**
3.1 XML - eXtensible Markup Language . . . . . 3
3.2 SOAP - Simple Object Access Protocol . . . . . 3
3.3 WSDL - Web Service Description Language . . . . . 5
3.4 UDDI - Universal Description, Discovery and Integration . . . . . 7
3.5 Zusammenfassung . . . . . 8

**4 Sicherheitsaspekte** **9**

**5 Anwendungsbeispiele** **9**
5.1 B2B - Business to Business . . . . . 9
5.2 EAI - Enterprise Application Integration . . . . . 9

# 1 Motivation

Durch den Boom des Internets während der 90iger Jahre und die weiterhin steigende Anzahl an neuen Informationen, ist es notwendig geworden, möglichst viele Bereiche des Internets zu automatisieren. So ist es also ein Ziel, Anwendungen miteinander kommunizieren zu lassen, ohne das der Datenaustausch “von Hand“ geregelt werden muss. Hierfür sind natürlich gewisse Voraussetzungen notwendig. Ein Ansatz wurde bereits 1991 von der *Object Management Group (OMG)* entwickelt. Bei *CORBA (Common Object Request Broker Architecture)* wird durch Einsatz von Middleware eine orts-, plattform- und implementations-unabhängige Kommunikation zwischen Applikationen ermöglicht.
Die Idee der Web Services besteht hingegen darin, durch den Einsatz von webbasierten Standards, Middleware zu ersetzen. Wie könnte nun also so ein Web Service aussehen? Ein beliebtes Beispiel hierfür ist die Reiseplanung. Man reserviert in Internet einen Flug und möchte nun auch gleichzeitig das Hotel, einen Mietwagen, etc. buchen. Bisher musste man sich nun die entsprechenden Seiten suchen und dann die Termine abgleichen. Web Services ermöglichen nun, dass durch das System der Fluggesellschaft die nötigen Informationen weitergegeben werden und man direkt Angebote der passenden Hotels, ... erhält.

# 2 Definition

*A Web Service is a software application identified by a URI, whose interfaces and binding are capable of being defined, described and discovered by XML artifacts and supports direct interactions with other software applications using XML based messages via Internet-based protocols.*[1]

Ein Web Service ist folglich eine Softwareanwendung, welche durch XML definiert wird und mittels bestehender Internet-Protokolle (TCP/IP, ...) mit anderen Anwendungen kommunizieren kann.
Web Services lassen sich über einen Protokollstack beschreiben. Die in Abb. 1 (S. 2) bezeichneten Schichten übernehmen hierbei folgende Aufgaben. Als Basis dient die Transportschicht. Hier wird das verwendete Transportprotokoll festgelegt. Die darüberliegende Schicht ist dafür verantwortlich, die Nachrichten in ein XML-Format umzuwandeln. Diese Schicht wird durch SOAP repräsentiert. Die Funktionalität des Dienstes wird in der nächsten Schicht (*Description*) definiert. Zuletzt wird dann die Entdeckung des Dienstes durch UDDI geregelt.
Hieraus lässt sich bereits die Bedeutung der Standards SOAP, WSDL und UDDI, erkennen, welche im folgenden noch genauer beschrieben werden.

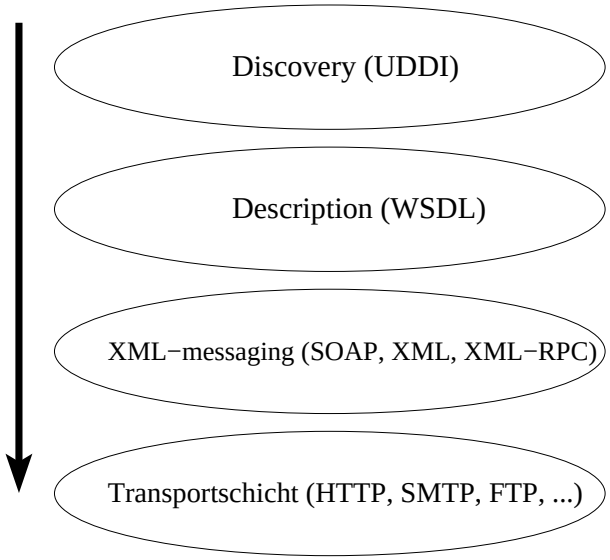



Abbildung 1: Protokollstack

Alternativ lassen sich Web Services auch durch die Aufteilung in verschiedene Rollen beschreiben. Dabei exisitieren zumeist drei Rollen.

- Dienstanbieter (Service Provider)
  - Bietet Dienstleistung über das Internet an
  - Beschreibt und annonciert diesen Dienst
  - Übernimmt Implementierung, Wartung und Betrieb
- Dienstnachfrager (Service Requestor)
  - Nutzt angebotenen Dienst
  - Sucht angebotene Dienste nach gewissen Kriterien
  - Integriert Service in eigene Dienste und Applikationen
- Dienstmakler (Service Broker)
  - Bietet Speicherung von Dienstbeschreibungen an
  - Verwaltet kategorisierte Dienstbeschreibungen
  - Ermöglicht automatisiertes Auffinden existierender Dienste

Da diese drei Rollen in der Regel verteilte Komponenten sind, können keine identischen Grundvoraussetzungen gemacht werden. Deshalb soll durch die Einführung von Standards die Unabhängigkeit der Dienste von sowohl der Plattform, als auch der zugrundeliegenden Programmiersprache erreicht werden. In Verbindung mit der weiten Verbreitung (speziell der guten Verfügbarkeit) des Internets ist somit die Integration eines Dienstes relativ einfach. Die daraus resultierende lose Kopplung ermöglicht das dynamische Binden der jeweiligen Dienste zur Laufzeit. Durch den Einsatz vom XML sind Web Services stark dokumentenorientiert.

# 3 XML & Co.

## 3.1 XML - eXtensible Markup Language

Wie bereits erwähnt, bauen Web Services auf XML auf. Aber was ist eigentlich XML?

XML ist eine Metasprache, d.h. sie erfasst nicht nur die formalen Strukturen des Textes, sondern sorgt dafür, dass Dokumente in ihrem Aufbau gewissen Grundmustern folgen. Durch Trennung von Form und Inhalt sind Programmierer in der Lage, Anwendungen zu schreiben, welche mit solchen Dokumenten arbeiten können. Genauso wie HTML ist auch XML ein textbasiertes Format, welches keine spezielle Anforderungen an Hardware oder Netzinfrastruktur stellt.

## 3.2 SOAP - Simple Object Access Protocol

Bei der Bereitstellung eines Diesntes über das Internet müssen aus Sicht des Dienstanbieters folgende Probleme gelöst werden:

- Abwicklung der Kommunikation (Protokoll)
  - synchrone Aufrufe über HTTP: Aufrufer wartet auf Antwort
  - asynchrone Aufrufe über SMTP: Entkopplung von Aufruf und Antwort
- Struktur des Kommunikationsinhalts

Die bekannteste Lösung hierfür stellt SOAP dar. Die Grundidee dabei ist es, entfernte Methodenaufrufe und Nachrichtenaustausch zu ermöglichen. Dabei trennt SOAP zwischen Nutzinformationen und Metadaten und bietet eine XML-Darstellung beliebiger Inhalte. Wie aus Abb. 1 (S. 2) leicht ersichtlich ist, ist SOAP unabängig von dem gewählten Transportprotokoll.
Abbildung 2 (S. 4) zeigt, wie eine SOAP Nachricht aufgebaut ist.

- Protocol Header
  Der Protocol Header ist von dem Protokoll (z.B. HTTP) abhängig, auf dem die SOAP Nachricht transportiert wird. Darin befindet unter anderem auch Angaben zur Übertragungsmethode (POST/GET) und zum Zielrechner.

- SOAP Envelope
  Die SOAP Envelope ist ein Container für die beiden Elemente SOAP Header und SOAP Body. Der SOAP Header enthält Informationen über die SOAP Nachricht, während der SOAP Body die eigentliche Nachricht enthält.

- SOAP Header
  Wie oben erwähnt sind im Header Informationen über die SOAP-Nachricht enthalten. Diese Informationen sind jedoch optional. Ebenfalls im Header können Anweisungen für die aufgerufene Anwendung enthalten sein.

- SOAP Body
  Der SOAP Body enthält die eigentliche XML-Nachricht. In ihr ist das aufzurufende Objekt mit der entsprechend aufzurufenden Methode angegeben. Weiter enthält der Body die zu übergebenden Parameter als einzelne Tags.

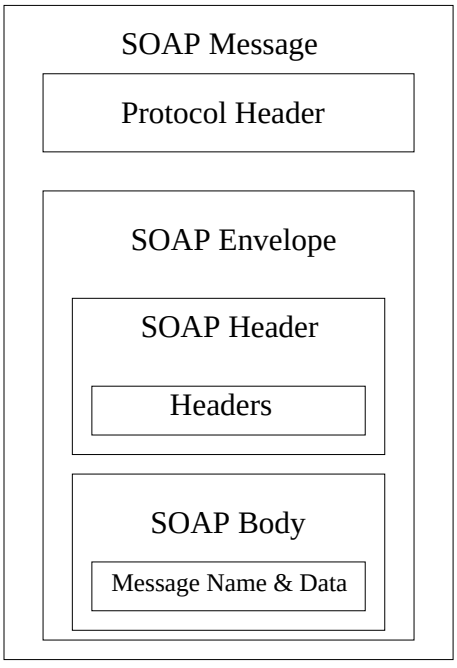



Abbildung 2: Aufbau einer SOAP Nachricht

Programm 3.1 (S. 5) ist ein Beispiel für eine SOAP Nachricht. Hier wird unter der Adresse *http://www.stock.org/stock* die Methode *GetStockPrice* aufgerufen. In diesem Fall bekommt die Methode den Namen der Aktie (*IBM*) als Parameter übergeben.

**Programm 3.1** Beispiel für einen Aufruf mit SOAP

```
<soap:Envelope>
 <soap:Body>
  <xmlns:m="http://www.stock.org/stock" />
   <m:GetStockPrice>
   <m:StockName>IBM</m:StockName>
  </m:GetStockPrice>
 </soap:Body>
</soap:Envelope>
```

Eine Antwort auf diese Anfrage könnte folgende Form haben (Programm 3.2, S. 5).

**Programm 3.2** Beispiel für eine Antwort mit SOAP

```
<soap:Envelope>
 <soap:Body>
  <xmlns:m="http://www.stock.org/stock" />
   <m:GetStockPriceResponse>
   <m:Price>34.5</m:Price>
  </m:GetStockPriceResponse>
 </soap:Body>
</soap:Envelope>
```

## 3.3 WSDL - Web Service Description Language

Mit Hilfe von WSDL wird definiert, welche Methoden beim Dienstanbieter aufgerufen werden können und welche Parameter diese benötigen. Desweiteren wird hier festgelegt, welche Rückgabewerte diese Methoden produzieren. WSDL Dokumente sind dabei in zwei Teile aufgeteilt (siehe Abb. 3, S. 6). Im ersten werden abstrakt die verwendeten Datentypen (*types*), die übertragenen Daten (*messages*) und die aufrufbaren Methoden, sowie deren Parameter (*portType*) beschrieben. Im zweiten Teil werden dann konkret sowohl der Kommunikationsendpunkt (*service*), d.h. die Adresse (üblicherweise URL), als auch das verwendete Protokoll (*binding*) definiert. Ein WSDL

Dokument wird vom Dienstanbieter erstellt und über einen Dienstmakler dem Dienstnachfrager zur Verfügung gestellt (vgl. Abb. 4, S. 8). Wie SOAP, so ist auch WSDL im Grunde ein XML Dokument.
Das *Operation* Element (siehe Abb. 3, S. 6) kann dabei folgendes sein:

- "One-Way"
  Eine Nachricht, die der Dienstnachfrager an den Dienstanbieter sendet, auf die er aber keine Antwort erhalten wird
- "Request/Response"
  Eine Nachricht, die der Dienstnachfrager an den Dienstanbieter sendet, worauf dieser eine Antwort sendet
- "Solicit-Response"
  Eine Nachricht, die der Dienstanbieter sendet, worauf der Dienstnachfrager antwortet
- "Notification"
  Der Dienstanbieter sendet eine Nachricht an den Dienstnachfrager

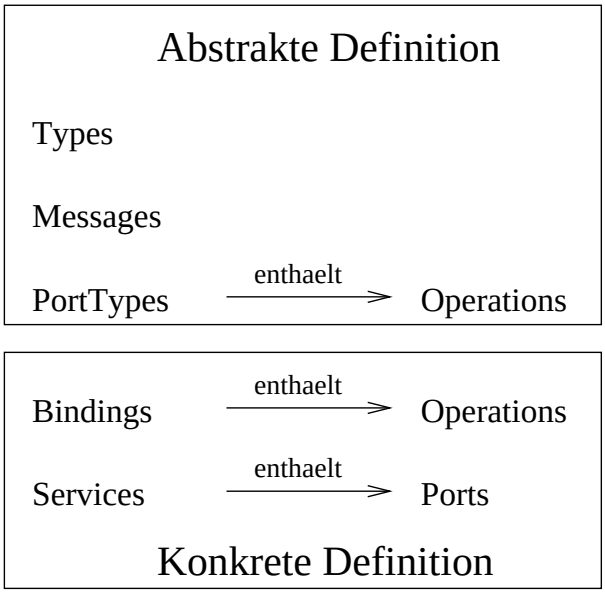



Abbildung 3: Aufbau eines WSDL Dokuments

Mit dem Element *ports* wird lediglich der Registrierungsname des Web Services angegeben.

## 3.4 UDDI - Universal Description, Discovery and Integration

UDDI wurde aus der Idee heraus entwickelt, ein globales, plattformunabhängiges Netzwerk zu erschaffen, um:

- Informationen über das Unternehmen in einer globalen Registry ablegen zu können
- Geschäftspartner zu ermitteln
- Geschäftsprozesse über Unternehmensgrenzen hinweg zu beschreiben und nutzbar zu machen

UDDI unterstützt die Suche nach dem passenden Dienst und bietet darüber hinaus die Möglichkeit, aus einer Großzahl von Unternehmen und Angeboten wählen zu können bzw. gewählt zu werden. Dadurch lassen sich die grundlegenden Fragen einer Geschäftsbeziehung

- Welche Firma den Dienst anbietet
- Wo sich der angebotene Dienst befindet
- Welcher Service angeboten wird
- Wie der Dienst benutzt werden kann

beantworten. UDDI stellt somit einen Verzeichnisdienst (und dadurch auch selbst einen Web Service) dar. Die hier verwalteten Daten werden kategorisch gespeichert und können gezielt gesucht werden. Die gängisten Kategorien sind

- “white pages“
  nach Namen sortiertes Register
- “yellow pages“
  nach Kategorien sortiertes Register
- “green pages“
  Detailinformationen

Ins Leben gerufen wurde UDDI von Ariba, IBM und Mircosoft, ist aber inzwischen ein weithin akzeptierter Standard.

## 3.5 Zusammenfassung

Abbildung 4 (S. 8) soll nun verdeutlichen, wie SOAP, WSDL und UDDI eingesetzt werden, um die Kommunikation zwischen Dienstanbieter, Dienstnachfrager und Dienstmakler zu regeln.

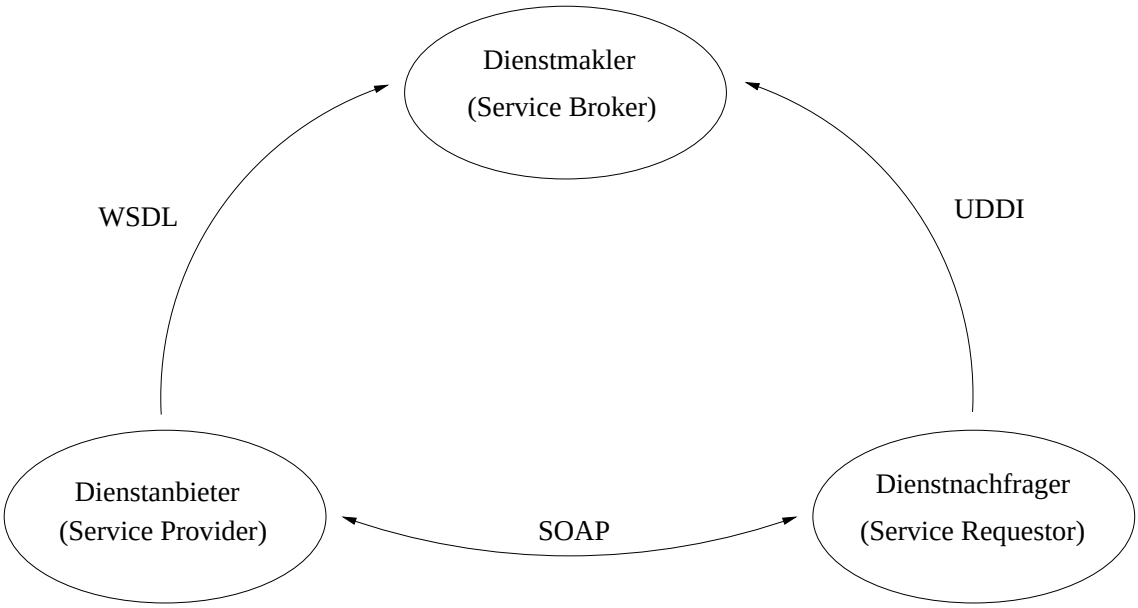



Abbildung 4: Web Service Architektur

# 4 Sicherheitsaspekte

Bei allen Anwendungen, welche das Internet als Medium benutzen, muss man sich die Frage stellen, ob sie ausreichende Sicherheitsmechanismen stellen. Gerade im Bereich B2B (Business to Business) müssen Authentifizierung, Authorisierung und Datenschutz sichergestellt sein. Genügt also SOAP solchen Anforderungen? SOAP bietet keine Mechanismen, welche den sicherheitstechnischen Anforderungen genügen würden. Aber das war auch kein Ziel für SOAP. Dennoch lassen sich die gewünschten Eigenschaften natürlich erreichen. Zum einen kann SOAP derart erweitert werden, dass durchaus akzeptable Resultate erzielt werden (vergleichbar mit dem SSL[1] -Zertifizierungsverfahren)[8]. Zum anderen lassen sich auch die existierenden Mechanismen des gewählten Transportprotokolls nutzen (beispielsweise HTTPS).

# 5 Anwendungsbeispiele

## 5.1 B2B - Business to Business

Wie bereits erwähnt lassen sich durch Web Services eigene Dienste fremden Unternehmen zur Verfügung zu stellen (z.B. Kreditkartenvalidisierung). Diese Art der Verwendung ist wohl die naheliegenste. Denkbare wäre aber auch die Realisierung unternehmensübergreifender Geschäftsprozesse mit Hilfe solcher Dienste. Somit können Lieferanten oder Partner in die eigene Systemwelt integriert werden.

## 5.2 EAI - Enterprise Application Integration

Eine weitere Einsatzmöglichkeit ist die Integration unternehmensinterner Geschäftprozesse. Dadurch können Kosten und Aufwand minimiert werden. Die Wiederverwertbarkeit der Dienste führt ebenfalls zu einer Kosteneinsparung.

[1] Secure Sockets Layer

# Literatur

[1] http://www.w3.org

[2] http://www.jeckle.de

[3] http://www.fzi.de

[4] http://www.firstsurf.com

[5] http://informatik.hsr.ch

[6] www3.informatik.tu-muenchen.de

[7] http://www.newtelligence.com/

[8] http://discuss.develop.com/

# Abbildungsverzeichnis

1 Protokollstack . . . . . 2
2 Aufbau einer SOAP Nachricht . . . . . 4
3 Aufbau eines WSDL Dokuments . . . . . 6
4 Web Service Architektur . . . . . 8

# Beispielprogramme

3.1 Beispiel für einen Aufruf mit SOAP . . . . . 5
3.2 Beispiel für eine Antwort mit SOAP . . . . . 5

# Web-Controlling

## Controlling-Kennzahlen für Websites

Seminar Internet-Dienste SS2003
Betreuer: Thilo Maier
Fakultät für Mathematik und Wirtschaftswissenschaften
Universität Ulm

# Inhaltsverzeichnis:


**1 Einführung ..... 3**

**2 Web-Controlling ..... 4**

**3 Kennzahlen des Web-Controlling ..... 4**

**3.1 Basiskennzahlen aus Rohdaten der Website-Nutzung ..... 5**

**3.2 Kombination von Rohdaten ..... 10**

**3.3 Kennzahlen mit Zeitbezug ..... 12**

**3.4 Kennzahlen zur Analyse der Kundenbeziehung ..... 14**

***3.4.1 Kennzahlen entlang des „Customer-Life-Cycle“ ..... 14***
***3.4.2 MikroKonversionsraten ..... 18***
***3.4.3 Recency, Frequency und Monetary Value (RFM) ..... 20***

**3.5 Kennzahlen zur Analyse des Kundenverhaltens einer Website ..... 22**

**3.6 Kennzahlen zur Erfolgsanalyse von Websites ..... 24**

**3.7 Komplexe Kennzahlen: Personalisierungsindex, Life Time Value, Loyality Value, Freshness Factor ..... 25**

**4 Fazit ..... 27**

**Literaturverzeichnis: ..... 28**

# 1 Einführung

„E-Business wird wieder Business!“, so überschreibt IBM in einer Werbung das „zweite Kapitel“ des Electronic Business. Die Anspielung bezieht sich auf die Euphorie, mit der das „erste Kapitel“ angegangen wurde, d.h. die Dotcom-Hysterie, in der man glaubte, die Regeln der Ökonomie außer Kraft setzten zu können. [Ringlein: S.283]

Verursacht wurde diese Euphorie durch die zahlreich hervorgesprossenen Internetfirmen und Start-Ups, deren Geschäftsmodelle auf der beinahe vollständigen Abwicklung aller Prozesse im Internet basierten und zu einem schnellen Erfolg führten. Viele Unternehmen konnten ihr Geschäftsmodell allerdings nicht lange erfolgreich umsetzen, da grundlegende betriebswirtschaftliche Tools einfach nicht angewandt wurden. Während es in den Anfangszeiten des E-Business Zeitalters lediglich wichtig erschien, die Zugriffe auf die eigene Website zu maximieren und dadurch den Bekanntheitsgrad des Unternehmens zu erhöhen, besinnen sich die Unternehmen heute wieder auf bewährte Mittel einer rationalen Unternehmensführung. Budgetverantwortliche sind nicht mehr bereit, Geld in Maßnahmen zu stecken, deren Return-on-Investment eher zweifelhaft ist. Kein traditionelles Unternehmen würde so arbeiten.

Die Fragen, die daraus resultieren, sind denkbar einfach:

1. Wie erfolgreich ist die Website und die Promotion dieser Site - gemessen an den Unternehmenszielen?
2. Was kostet das?
3. Rechtfertigen die Ergebnisse die aufgewendeten Summen?

Zur Beantwortung dieser Fragen werden Web-Controlling-Systeme, die Aussagen über die Kosten und den Nutzen des Internetauftritts erlauben, immer populärer.

# 2 Web-Controlling

Im Fokus des Web-Controlling stehen eher die neuen Herausforderungen an Online-Geschäftstätigkeiten, die durch die Website-Nutzung entstanden sind, als die traditionellen Aufgabenfelder des Controlling. Das Web-Controlling beschreibt das Umfeld derjenigen Aktivitäten, mit denen man die Ausgestaltung des Web-Auftritts von Unternehmen überprüft und Hinweise gibt, wo und wie Verbesserungsmöglichkeiten umgesetzt werden können. [Hippner, S.236]

Es bietet demnach Lösungen zur Planung, Analyse und Steuerung sämtlicher Web-Aktivitäten und eröffnet die Möglichkeit, Daten für das Planen und Optimieren der gesamten E-Business-Aktivitäten zu sammeln und zu bewerten. Vom Klick auf das Werbebanner über die Site in den Shop und sogar bis in das Warenwirtschaftssystem kann erstmals das gesamte Verhalten der Kunden erfasst und analysiert werden.

Dabei spielen Kennzahlen eine wichtige Rolle. Sie dienen dazu, schnell und prägnant über ein ökonomisches Aufgabenfeld zu berichten und bilden somit das Rückgrat für eine ergebnisorientierte Steuerung. Das Web-Controlling beschäftigt sich also mit der Sammlung und Auswertung von Kunden- und Nutzungsdaten, die dann zu aussagekräftigen Kennzahlen verdichtet werden.

# 3 Kennzahlen des Web-Controlling

Bisherige Erfolgskontrollen zu Webaktivitäten der Unternehmen werden derzeit im Wesentlichen aus reiner Marketingsicht betrieben. So wird aus dieser Sicht das Verhalten der Besucher auf den Websites des Unternehmens analysiert. Hierbei lassen sich vor allem folgende Analysekriterien zuordnen:

- Seitenanalysen nach Page Views, Visits und Verweildauer

- Besuchsanalysen nach Herkunft, Firma, Organisation und Verhalten auf Websites
- Navigationsanalysen zu Click-Pfaden innerhalb des Webangebots, Anzahl abgerufener Seiten und typische Ein- und Ausstiegsseiten. [Müller:S.2-3]

Die Auswertungen werden auf der Basis automatisch erstellter Logfiles generiert. Logfiles sind Protokolldateien, in denen automatisch bestimmte Abläufe der Rechnertätigkeit festgehalten werden, die während eines Nutzenvorgangs (Session) vom Webserver aufgezeichnet werden. Sie enthalten im Allgemeinen Informationen über den Zugriffszeitpunkt, die angeforderten Inhalte, den zugreifenden Browser, die Adresse des Client, eventuell auftretende Fehler und die URL der Seite, die den Link auf die angefragte Seite enthielt (Referrer).
Aus den Logfiles generierten Basisdaten lassen sich durch spezielle Analyseprogramme aussagefähige Controlling-Kennzahlen erstellen.

### 3.1 Basiskennzahlen aus Rohdaten der Website-Nutzung

Weitverbreitete Basiskennzahlen zur Messung des Nutzerverhaltens auf Websites bilden die Messgrößen Hits, Page Views, Visits und User. [Schwickert: S.4]
Ein User ist hierbei ein Nutzer, der auf eine Website zugreift. Ein Visit (auch Session) beschreibt einen zeitlich zusammenhängenden Nutzungsvorgang eines bestimmten Users, wobei ein Visit aus einem oder mehreren Page Views bestehen kann. Als Page View (auch Page Impression) wird der Sichtkontakt eines beliebigen Besuchers mit einer vollständigen Website bezeichnet. Dabei kann eine Seite mehrere Objekte beinhalten, die als eigenständige Dateien auf dem Server vorliegen. Jede Anfrage an den Server, und damit jede übertragene Datei stellt einen Hit dar. Ein Hit wird in der Logfile-Datei durch einen Eintrag (Zeile) repräsentiert [Hippner, S.132]. Vgl. Abbildung **3-1**

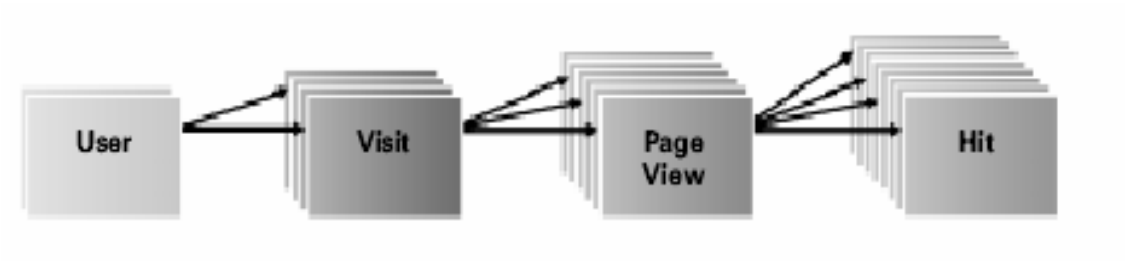



Abbildung **3-1**: Hierarchie der Website Aktivität [NetGenesis, S.16]

Grundsätzlich sind die Website-Anbieter an möglichst hohen, im Zeitverlauf steigenden Zugriffszahlen interessiert, da eine hohe Zahl an Besuchern und Nutzern auch eine entsprechende Anzahl an (potenziellen) Kunden nahelegt. Eine hohe Anzahl von Hits pro Besuch belegt allerdings nicht unbedingt, dass es sich um ein interessantes Angebot handelt, sondern kann durch eine umständliche Seiten-navigation verursacht werden. Insbesondere gilt dies für Online-Shops, da der direkte Weg zum Ziel einen wesentlichen Vorteil für den Kaufinteressenten darstellt. Angesichts dieser Defizite sind Hitstatistiken zur Einschätzung des Erfolges von Websites nur bedingt geeignet.

Auch die Anzahl der Page Views wird durch lange Seitenpfade, z.B. beim Ausfüllen eines Bestellformulars in mehreren Schritten durch die Gestaltung der Seiten-abfolge nach oben verfälscht. Ebenso werden vor allem im Servicebereich niedrige Nutzungszahlen angestrebt, denn somit kann man davon ausgehen, dass die Kunden zufrieden sind. Vergleichsweise hohe Zahlen sind dagegen bei Produktinformationen und Online-Bestellungen wünschenswert. Deshalb sollten Page Views unbedingt getrennt nach den Bereichen einer Website ausgewertet werden, um Anhaltspunkte darüber zu bekommen, welche Seitenbereiche überarbeitet oder hervorgehoben werden müssen und welche Seitenbereiche zufriedenstellende Nutzungszahlen aufweisen. [Schwickert: S.4]

Aussagekräftiger, aber aufgrund technischer Gegebenheiten schwieriger zu messen, sind Visits und die Anzahl der Nutzer. Um zusammenhängende Nutzungsvorgänge von Websites erfassen zu können, muss dem jeweiligen Besucher für die

Dauer der Session mit der angeforderten Seite vom Server eine eindeutige Kennung (Session-ID) zugewiesen werden. Dadurch lässt sich der Bewegungspfad eines Nutzers rekonstruieren und die Anonymität des Nutzers bleibt gewahrt. Allerdings geht bei einem erneuten Besuch auf der Website die Session-ID verloren, und dem Besucher wird eine neue Kennung zugeordnet. Oft wird auch zur schärferen Abgrenzung einer Session verlangt, dass die Zeitspanne zwischen zwei Seitenaufrufen einen bestimmten Maximalwert (z.B. 30 Minuten) nicht überschreiten darf. Damit ist es nicht möglich, individuelle Nutzerprofile auf Basis von Session-IDs zu erstellen. [Hippner, S.18,111]

Die Berechnung von Kennzahlen auf Basis der Nutzeranzahl setzt die Identifikationsmöglichkeit einzelner Nutzer voraus, z.B. durch Cookies oder IP-Adressen. IP-Adressen sind jedoch nicht immer eindeutig, da sie oft dynamisch zugewiesen werden und somit dem Nutzer zu verschiedenen Zeitpunkten verschiedene Adressen zugeordnet werden oder wenn der Nutzer von verschiedenen Rechner aus auf das Internet zugreift. Auch besteht die Möglichkeit, dass mehrere Nutzer mit der gleichen IP-Adresse erfasst werden beim Zugang zum Internet über einen Proxy-Server oder wenn sie mit mehreren Nutzern ihren Rechner teilen. Selbst bei der Verwendung von Cookies wird lediglich ein bestimmter Rechner identifiziert. Wird dieser Rechner von mehreren Personen genutzt, kann die Zuordnung von Zugriffen zu einzelnen Personen nur noch über eine Registrierung (Benutzername und Passwort) erreicht werden. [Hippner, S.17]

Eine wichtige Kennzahl liefert hier die Anzahl der registrierten Nutzer. Mit der Sammlung dieser Daten erhält der Website-Anbieter eine Liste von Personen, die bspw. per Email gezielt auf ihre Bedürfnisse angesprochen werden können. Durch eine detaillierte Betrachtung der Nutzeranzahl können Nutzerprofile erstellt werden, die wiederum zur Identifikation erfolgsversprechender Kundenpotenziale verwendet werden können.

Durch die Logfileauswertung lässt sich auch der Weg eines Besuchers beim Navigieren durch die Website, der sogenannte Clickstream nachvollziehen. Dieser gibt Aufschluss darüber, welche Angebote genutzt werden und wo die häufigsten Ein- und Ausstiegsseiten sind. Während eine abgeschlossene Bestellung eines Online-Shops als Ausstiegsseite sicher positiv zu bewerten ist, muss bei einem Abbruch auf der Homepage das Seitendesign und die inhaltliche Gestaltung hinterfragt werden.

Aus Marketingsicht ist besonders interessant, wer die Website benutzt, wie die Nutzer zur Website gelangen, über welche Seite sie das Angebot wieder verlassen und welche Aktionen sie in der Zwischenzeit ausführen bzw. welche Pfade sie durch die Website nehmen. Die Referrer-Einträge in den Logfiles geben Auskunft darüber, ob der Benutzer durch einen Klick auf Werbebanner (Click-Through oder auch AdClick), durch einen normalen Link oder durch eine Suchmaschine auf das eigene Angebot überwiesen wurde. Wenn das Unternehmen auf einer anderen Website derartige Banner einsetzt, ist es für die Bestimmung der Effizienz dieser Werbung notwendig, die Anzahl der aktiven Kontakte zu messen sowie die Reichweite, die angibt, wie oft ein Banner auf einer Seite geschaltet wurde. Wird die Website über eine Suchmaschine erreicht, wird nicht nur deutlich über welche Suchmaschine der Nutzer auf die Seite gelangt, sondern auch durch welchen Suchbegriff. Eine Auswertung der häufigsten Suchbegriffe ermöglicht gezielte Eintragungen der Website bei den Suchmaschinen, die dazu führen sollen, eine höhere Anzahl und präzisere Auswahl neuer Benutzer auf die Website aufmerksam zu machen.

Auch ist es möglich, über den Domain-Namen jene Länder herauszufinden, aus denen die Nutzer kommen bzw. welchen Organisationen sie angehören. Wird die Website in größerem Umfang auch von fremdsprachlichen Nutzern besucht, sollte man darüber nachdenken, den Inhalt auch in anderen Sprachen anzubieten. Unter den Domain-Namen sind teilweise auch Namen von Wettbewerbern zu entdecken. Durch geeignete Filterung nach diesem Kriterium wird verdeutlicht, welche Seitenbereiche durch Wettbewerber besonders beobachtet werden und liefert damit einen Beitrag zur Wettbewerbsanalyse.

Bezeichnungen der Subdomains weisen oft auf Organisationen hin und enthalten z.B. Namen von Unternehmen, Universitäten oder Zugangsprovidern. Mit Hilfe dieser Information kann man den Inhalt noch besser an seine Benutzergruppe anpassen.

Sinnvoll ist der Einsatz von Auswertungssoftware wenn es um das Aufspüren von Fehlern in der Website geht, z.B. bei Nutzerzugriffen auf nicht mehr vorhandene Seiten. Schreibfehler und Links, die sich auf veraltete Adressen beziehen sind die häufigsten Ursachen und lassen sich teilweise durch geeignete Gegenmaßnahmen beheben. Die Website sollte außerdem dahingehend überprüft werden, wie sie von den am häufigsten genutzten Browsern und Betriebssystemen angezeigt wird. Ältere Versionen sind unter Umständen nicht fähig, bestimmte Gestaltungselemente zu verarbeiten.
Abbildung **3-2** fasst die wichtigsten Kenngrößen nochmals zusammen:

| **Kenngrößen aus Rohdaten der Website-Nutzung** | | |
|---|---|---|
| Hits | Registrierte Nutzer | Suchbegriffe |
| Page Views | Nutzungszahlen von Seiten | Top-Level-Domains |
| Visits | Einstiegsseiten | Subdomains |
| User | Ausstiegsseiten | Üertragenes Datenvolumen |
| Ein-Klick-Besuche | Linknutzung | Üertragungsfehler |
| Weitergehende Besuche | Reichweite Bannerwerbung | Browser-Typen,-Versionen |
| Mehrfachbesucher | Click-Throughs der Banner | Betriebssysteme |

Abbildung **3-2**: Kenngrößen aus Rohdaten der Website-Nutzung [Schwickert, S.8]

## 3.2 Kombination von Rohdaten

Verwendet man ausschließlich Kennzahlen auf der Basis von Rohdaten, ist die Analyse der Website wenig aussagekräftig, da den Werten Vergleichsmaßstäbe fehlen. Verhältniszahlen, die die absoluten Zahlen in Form von prozentualen Anteilswerten wiedergeben, relativieren die Angaben über die Nutzer und deren Besuche.

Eine wichtige Größe in diesem Zusammenhang ist das Verhältnis zwischen der Anzahl der Page Impressions und Visits. Dieses sagt einiges über das Nutzungsverhalten der User aus, insbesondere über die Intensität der Nutzung. Ein hoher Prozentsatz an Ein-Klick-Besuchen deutet zwar auf effektive Promotion-Maßnahmen hin, die Nutzer auf die Website führen, doch die Attraktivität des präsentierten Inhalts scheint vergleichsweise niedrig zu sein. Website-Anbieter sind daher an einem hohen Anteil weitergehender Nutzungsvorgänge interessiert.

Aus den Daten der Logfiles lässt sich auch die Reihenfolge der Seiten bestimmen, die sich der Nutzer angesehen hat. Eine Analyse der meistgenutzten Klickpfade kann Hinweise auf das Nutzerverhalten geben, ob Bestellungen erst nach ausführlicher Recherche der Produktinformation folgen oder ob plötzliche Abbrüche der Nutzung in Verbindung mit Navigationsproblemen auftreten.

Aus der Kombination des Nutzerverhaltens mit demographischen und psychographischen Daten lassen sich detaillierte Nutzerprofile erstellen. Diese können als Grundlage für eine Personalisierung dienen. Allerdings müssen bei der Datennutzung die gesetzlichen Rahmenbedingungen eingehalten werden, um das Vertrauen der Besucher nicht aufs Spiel zu setzen. Der Anteil verwendeter Nutzerinformationen misst, inwieweit vorhandene Kundeninformationen - Namen, Geburtsdaten, Hobbies, bislang gekaufte Produkte usw. - auch wirklich genutzt werden. Ein hoher Anteil an verwendeten Nutzerinformationen muss nicht unbedingt auch einen hohen Personalisierungsgrad in der Ansprache bedeuten. Denn werden

nur zwei verwendete Kundeninformationen erfasst und beide genutzt, ist der Anteil zwar optimal, aber die zwei Elemente an sich sind für eine personalisierte Ansprache nicht hinreichend.

Um die Effektivität einer Werbemaßnahme zu beurteilen, ist es hilfreich, die Klickrate (Click-Through-Rate) zu bestimmen. Diese errechnet sich aus dem Verhältnis von Click-Throughs zur Anzahl der Abrufe einer werbetragenden Seite. Der Wert der Klickrate liegt in der Praxis meist unter 0,5% [Strauss: S.16]. Im Zusammenhang mit Fragebögen stellt sich immer die Frage wie repräsentativ die Ergebnisse sind. Erst die Antwortquote, also das Verhältnis ausgefüllter Fragebögen zur Gesamtzahl der Page Views der entsprechenden Seite, zeigt, wie erfolgreich eine Fragebogenaktion ist und wie sehr die Nutzer bereit sind, persönliche Informationen weiterzugeben. Allerdings muss den Nutzern hier in der Regel ein Anreiz gegeben werden, damit diese an den Umfragen teilnehmen, z.B. in Form von Gewinnspielen.

Abbildung **3-3** fasst die wichtigsten Kenngrößen zusammen:

| **Kenngrößen aus der Kombination von Rohdaten** | | |
|---|---|---|
| Anteil der Ein-Klick-Besuche | = | Ein-Klick-Besuche / Visits gesamt |
| Anteil weitergehende Besuche | = | Weitergehende Besuche / Visits gesamt |
| Besuchstiefe | = | Page Views / Visits |
| Anteil der Mehrfachnutzer | = | Mehrfachnutzer / User gesamt |
| Anteil der registrierten Nutzer | = | Registrierte Nutzer / Nutzer gesamt |
| Klickpfade | = | Reihenfolge der Seitenaufrufe |
| Click-Through-Rate | = | Click-Throughs / Reichweite |
| Antwortquote von Fragebögen | = | Ausgefüllte Fragebögen / Page Views der Seite |

Abbildung **3-3**: Kenngrößen aus der Kombination von Rohdaten [Schwickert, S.11]

### 3.3 Kennzahlen mit Zeitbezug

Kennzahlen mit Zeitbezug geben zum einen Auskunft darüber zu welcher Tageszeit oder an welchen Wochentagen die Website am stärksten frequentiert wird. Diese Information lässt sich leicht aus den Zeitangaben eines ausgewerteten Logfiles ablesen und kann für beliebige Zeiträume analysiert werden. Zum anderen können zeitbezogene Kennzahlen über die durchschnittliche Besuchsdauer oder die Dauer der Betrachtung einzelner Seiten oder Bereiche Auskunft geben. Wird das Nutzungsverhalten über längere Zeiträume hinweg beobachtet, lassen sich Prognosen über die zukünftige Nutzung ableiten. [Schwickert, S. 13]

Ein grundlegendes Zeitintervall zur Beschreibung der Website-Nutzung ist die Verweildauer. Die Verweildauer ist die Zeitspanne, die ein Nutzer auf einer bestimmten Website verbringt. Dabei wird die Zeit zwischen dem ersten und letzten Zugriff auf die Seite gemessen. Die Verweildauer gibt Auskunft darüber wie intensiv die Besucher das Angebot der Website nutzen. Wird diese Kennzahl im Zusammenhang mit dem Inhalt einer Seite betrachtet, kann festgestellt werden, ob die Information vom Besucher erfasst werden konnte oder nur flüchtig registriert wurde. Man darf nicht generell davon ausgehen, dass eine lange Verweildauer etwas Positives darstellt, z.B. sollten Seiten des Kundenservices nur kurz genutzt werden. Um zu messen wie lange ein Seitenbereich die Aufmerksamkeit der Nutzer auf sich zieht, wird die durchschnittliche Verweildauer aller Nutzer für die betreffende Seite ermittelt.
Die Besuchsfrequenz liefert eine Aussage über die regelmäßige Nutzung des Webangebots durch einzelne Besucher. Ist diese Zahl hoch, dann kann davon ausgegangen werden, dass das Informationsangebot auf der Website attraktiv ist. Für Service-Seiten sollten hier allerdings niedrige Werte angestrebt werden, da dies ein Zeichen für die Kundenzufriedenheit ist und das Unternehmen Kosten hinsichtlich der Bearbeitung der Anfragen einsparen kann. Neben der Besuchsfrequenz kann für jeden Besucher der Zeitraum ermittelt werden, in dem er der Website fernbleibt. Wie groß dieser Zeitraum üblicherweise ist, hängt wie bereits erwähnt vom Angebot der Website ab. Die Online-Version einer Tageszeitung hat einen regelmäßigen

Nutzer wahrscheinlich verloren, wenn dieser über eine Woche nicht auf das Angebot zugreift. Auf Service-Seiten trifft dies natürlich nicht zu.

Die Besuchsfrequenz stellt dar, wie oft ein Nutzer in einem bestimmten Zeitintervall auf eine Website zugreift. Dies ist wieder abhängig vom Angebot der Website. In der Regel wird sich der Kunde kurz vor einer Bestellung vermehrt über das Produkt informieren. Die Besuchsfrequenz steigt an und der Anbieter kann, sofern er dies rechtzeitig bemerkt, darauf reagieren und den potenziellen Kunden mit einem Anreiz endgültig überzeugen.

Abbildung **3-4** fasst die wichtigsten Kenngrößen zusammen:

| **Kenngrößen mit Zeitbezug** | | |
|---|---|---|
| Verweildauer eines Besuches | = | $\Sigma$ Verweildauern einzelner Page Views |
| Durchschnittliche Verweildauer | = | $\Sigma$ Verweildauern einzelner Nutzer / Nutzeranzahl |
| Dauer des Fernbleibens | = | Aktuelles Datum – Datum des letzten Besuchs |
| Besuchsfrequenz | = | Besuche eines Nutzers / Zeitraum |
| Durchschnittliche Server-Antwortzeit | = | Zeitraum zwischen Client-Request und Server-Response |
| Durchschnittliche Download-Zeiten | = | $\Sigma$ Übertragungszeit / Anzahl Downloads |
| Täglicher/Wöchentlicher/Monatlicher Nutzungsverlauf | = | Nutzerzahlen, Seitenaufrufe, Server-Auslastungen über bestimmte Zeiträume |
| Störungsrate | = | Anzahl Störungen / Nutzungsdauer System |
| Störungsintensität | = | Ausfallzeiten durch Störungen / Nutzungsdauer |

Abbildung **3-4**: Kenngrößen mit Zeitbezug [Schwickert, S.14]

## 3.4 Kennzahlen zur Analyse der Kundenbeziehung

Ein weiterer Ansatz zur Erfolgskontrolle lässt sich im Vertriebsbereich feststellen. So versuchen Anbieter und Betreiber von Shoppinglösungen verstärkt das Verhalten der Nutzer im Online-Shop zu erfassen und systematisch auszuwerten. Auch hier liegt, im Gegensatz zum realen physikalischen Shop, eine umfassende Datenbasis vor. Der Nutzer hat bei seinen Aktionen im virtuellen Shop etliche digitale Spuren hinterlassen. Bei diesen Analysen wird versucht, durch die Verknüpfung der Marketing- und Besucherdaten der Website-Nutzung mit den Informationen aus der elektronischen Shopumgebung betriebswirtschaftliche Erfolgsgrößen der Online-präsenz zu ermitteln. Hierzu gehören:

- Produktanalysen nach Kategorien, Umsätzen und Preisen
- Warenkorbanalysen nach häufig gewählten Produkten und Produktgruppen sowie Cross-Selling-Aspekten
- Konversionsratenanalysen (vgl. 3.4.1) von der Website zur Shopansicht, von der Produktauswahl in den Warenkorb und vom Warenkorb bis zur Bestellung [Müller, S.3]

Da im Fokus des beziehungsorientierten Marketings der Aufbau und die Festigung langfristig profitabler Kundenbeziehungen stehen, kann der Erfolg an den Verände-rungen einerseits durch eine detaillierte Überwachung des Kaufverhaltens entlang des Kundenlebenszyklus (Customer-Life-Cycle) sowie bestimmter Produktpräferen-zen oder hinsichtlich der Frage nach dem „besten“ Kunden gemessen werden.

### 3.4.1 Kennzahlen entlang des „Customer-Life-Cycle“

Abbildung **3-5** liefert eine Übersicht des Vorgangs, aus potenziellen Kontakten im Zusammenhang mit Online-Shops einen treuen Kundenstamm aufzubauen und zeigt Ansatzmöglichkeiten für Web-Controlling Maßnahmen auf.

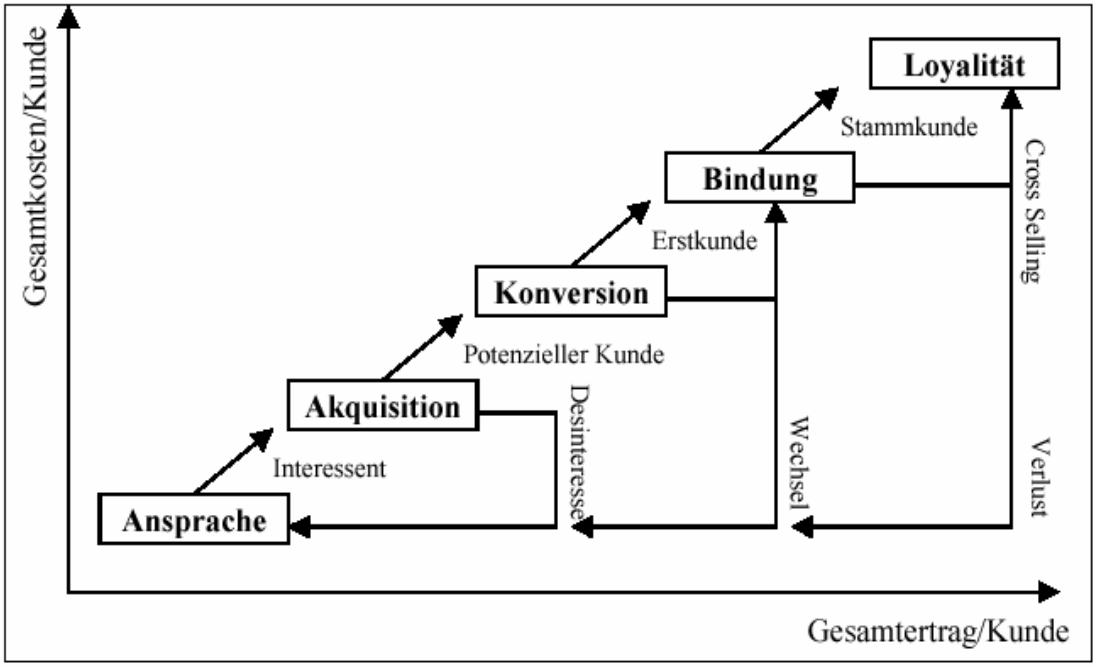



Abbildung **3-5**: Kundenlebenszyklus [Exody, S.34]

Der Kundenlebenszyklus kann in folgende Phasen unterteilt werden:

**Ansprache (Reach)**

In dieser Phase wird die Aufmerksamkeit eines Nutzers durch Werbe- oder Marketingmaßnahmen geweckt. Ziel dieser Phase ist es den Bekanntheitsgrad des Unternehmens zu steigern und möglichst ganze Marktsegmente zu erreichen.

**Akquisition (Acquisition)**

Ziel der Akquisition ist es, mit dem potenziellen Kunden in Kontakt zu kommen, um sein Kaufinteresse für ein bestimmtes Produkt zu wecken.

**Konversion (Conversion)**

Unter Konversion wird ein potenzieller Kunde verstanden, der Waren kauft und somit zu einem tatsächlichen Kunden wird.

**Kundenbindung (Retention)**

In dieser Phase wird versucht die Kundenzufriedenheit zu erhalten, in Kontakt zu bleiben und Wiederholungskäufe zu initiieren. Die Kosten, um einen bestehenden Käufer zum Wiederholungskauf zu bewegen, machen nur einen Bruchteil der durchschnittlichen Kosten von 150 bis 300 Dollar für die Neugewinnung von Kunden im Internet aus. Und wenn man bedenkt, dass Amazon jährlich einen durchschnittlichen Gewinn von 130 Dollar pro Kunde erzielt, so lässt sich leicht die Bedeutung der Kundenbindung erkennen. [Hippner, S.4]

**Loyalität (Loyalty)**

Diese Kunden haben den höchsten Wert, da sie ein starkes Markenbewusstsein entwickeln und eine hohe Kundenbindung aufweisen. Denn diese Kunden werden dem Unternehmen wahrscheinlich auch in Zukunft Umsätze verschaffen, ohne dass dabei besonderer Aufwand für Werbung oder Überzeugung anfällt. Anzeichen für eine hohe Loyalität sind eine hohe Besuchsfrequenz und eine hohe Wiederkaufrate. [Schwickert: S.15]

In jeder Phase des Kundenlebenszyklus besteht die Gefahr, dass der Kunde aus dem Zyklus aussteigt. Er kann direkt nach der Akquisition ausscheren, oder nach dem ersten Kauf zu einem anderen Anbieter wechseln. Eine Maßzahl für den zweiten Fall ist die sogenannte Churn-Rate. Sie misst wie viel bestehende Kunden im Verhältnis zur Gesamtkundenzahl den Anbieter wechseln.

Die Marketingaufwendungen, die vor allem aus den Kosten für den Internet-Auftritt des Unternehmens und seiner Website bestehen, können als Investition betrachtet werden, da die Zeitpunkte zwischen Kostenanfall und den zur Amortisation dieser Kosten beitragenden Umsatzerlöse sehr weit auseinander liegen können. Aus diesem Grund sind die Unternehmen bestrebt, dass die Kunden den Zyklus mög-

lichst schnell durchlaufen. Deshalb gehört zu den wichtigsten Kennzahlen des Zyklus der Return on Investment (ROI). Wenn Kunden z.B. mit einer 50.000 Euro teuren Bannerkampagne zu Kauf eines Produktes mit einer Marge von 500 Euro angeregt werden sollen, dann liegt der ROI bei 100 Käufern. Wie nah eine Werbekampagne am ROI-Punkt liegt, kann kontinuierlich gemessen werden.

Visualsierungstechniken wie der Kundenlebenszyklus enthüllen, wo es beim Durchlaufen des Zyklus „klemmt". Der Kundenlebenszyklus-Trichter visualisiert, wie viele Kunden von einer Stufe zur nächste vorschreiten oder aber abspringen. Verengt sich der Trichter an einer Stelle rapide, stimmt dort die „Sprungquote" nicht, und viele Kunden steigen aus dem Prozess aus. Die Idealform ist ein langsam enger werdender Trichter mit einer möglichst breiten, unteren Öffnung, vgl. Abbildung **3-6**.

Die Erhöhung der Akquisitions-, Konversions- und Kundenbindungsquoten – also die Optimierung der Trichterform – haben erheblichen Einfluss auf den ROI. Eine Möglichkeit, den Trichter der Idealform anzunähern, ist das Design des optimalen Klickpfades. Je nach seiner Position im Kundenlebenszyklus erhält ein Besucher verschiedene Angebote von Klickpfaden.

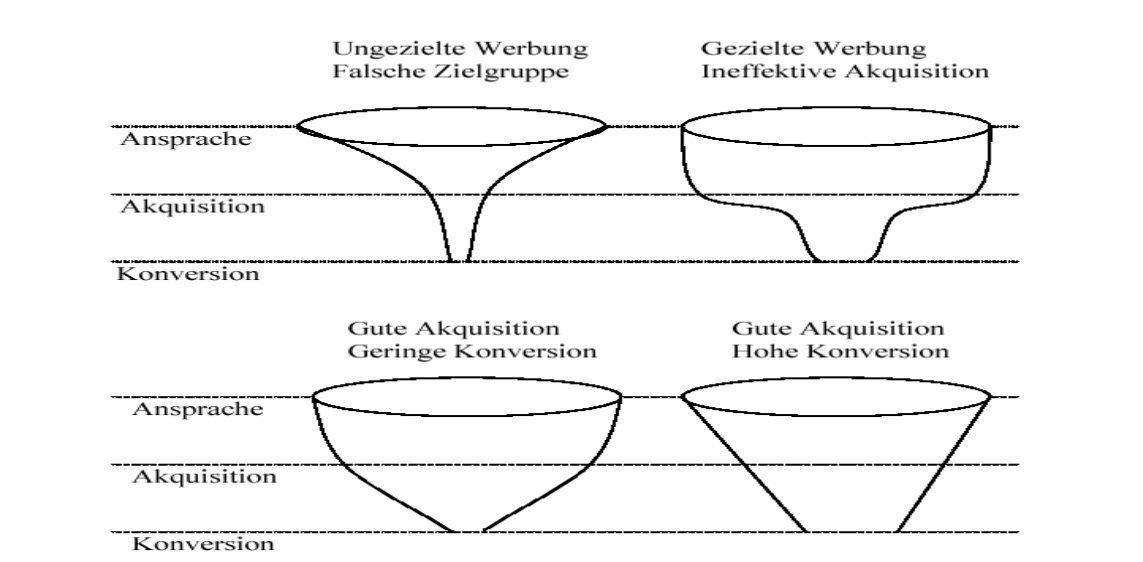



Abbildung **3-6:** Kundenlebenszyklustrichter [NetGenesis, S.45]

### 3.4.2 MikroKonversionsraten

Die traditionelle Konversionsrate eines Online-Shops misst den Anteil der Kunden, die tatsächlich Waren kaufen. Während diese Kennzahl nützlich ist, um die allgemeine Effizienz des Online-Shops zu messen, liefert sie keine Aussagen bezüglich der Faktoren, die die Verkaufsleistung beeinflussen. Die Mikrokon-versionsrate [vgl. Gomory, S. 4ff.] erweitert diese Kennzahl, in dem sie die folgenden Schritte des Wareneinkaufs mit einbezieht:

1. **look:** umschreibt die Phase, in der der Site-Besucher erste Eindrücke über das Online-Angebot sammelt (product impression)
2. **click:** umfasst den Vorgang, in dem auf einen speziellen Hyperlink geklickt wird und zugehörige Produktinformationen verfügbar werden (click-through)
3. **basket:** Situation, wenn ein Kaufobjekt in einen elektronischen Einkaufskorb gelegt wird (basket insertion)
4. **buy:** bezeichnet die finale Kauftransaktion (purchase)

Da sich nicht alle Kontaktmöglichkeiten realisieren lassen (siehe (1) Abbildung **3-7**), ergibt sich als erste Web-Controlling-Aktivität die Aufgabe zu überprüfen, wie der Anteil nicht realisierter Kontakte zu bewerten ist und evtl. zu verringern ist. Die Anzahl von (wo und wie) platzierten Werbebannern oder Angaben über die Präsenz und Positionierung des Web-Auftritts in Suchmaschinen liefern hier Beispiele für Informationen, die sich in entsprechende Web-Controlling Kennzahlen trans-formieren lassen, mit denen Erfolge bei der Kontaktaufnahme überprüfbar werden.

Gleiches gilt für die Ursachenforschung, warum Kontakte mit der Website nicht zum Kauf geführt haben (siehe (2) in Abbildung **3-7**). Möglicherweise liegen hier Schwachstellen in der Attraktivität der Website, um Nutzer zum Kauf eines Produktes zu überzeugen. Oder es werden Warenkörbe verlassen durch zu umfangreiche oder unverständliche Registrierprozeduren bei der Bestellung. Als drittes gilt

zu untersuchen warum Käufer nicht zu Wiederkäufern werden (siehe (3) Abbildung **3-7**).

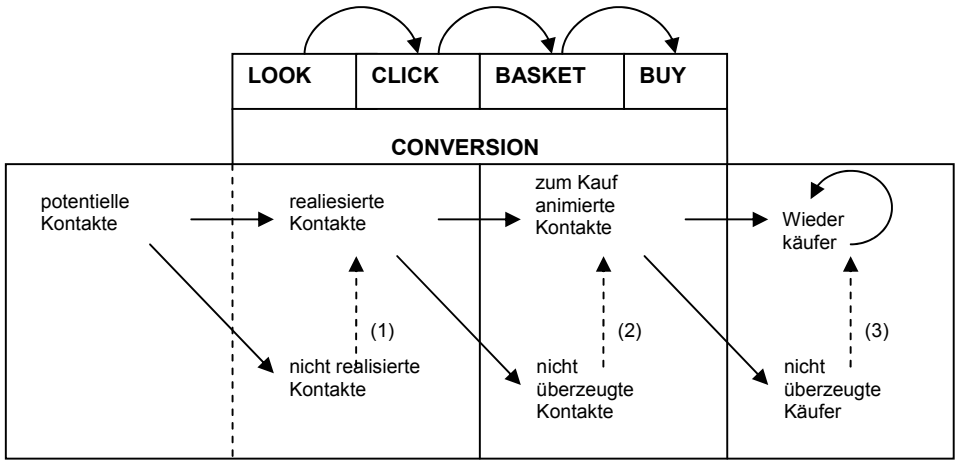



Abbildung **3-7**, Zustandswechsel bei Online – Geschäftsabläufen als Ansatzpunkte für Web-Controlling Aktivitäten [Hippner, S.238]

Wenn jeder Einkaufsschritt zueinander in Beziehung gesetzt wird (vgl. Abbildung **3-7**), können die drei folgenden Mikrokonversionsraten gebildet werden:

1. **look-to-click rate:** Prozentsatz der Besucher, die ein Angebot angeklickt haben, nachdem sie es gesehen haben
2. **click-to-basket rate:** Prozentsatz der Besucher, die danach ein Produkt in den Warenkorb gelegt haben
3. **basket-to-buy rate:** Prozentsatz der Besucher, die auch einen Kauf getätigt haben. Ist diese Rate gering entspricht sie den oberen zwei Kundenlebenszyklustrichtern vgl. Abbildung **3-6**

Durch diese genaue Überprüfung der Einkaufsschritte ist es möglich aufzuzeigen, auf welcher Ebene der Online-Shop die meisten Kunden verliert und wodurch dieser Verlust verursacht wird. Diese Mikrokonversionsraten lassen sich auch kombinieren, so dass der Prozentsatz der Käufer unter den Besuchern, die ein Produkt gesehen haben, berechnet werden kann, indem der erste und der letzte Einkaufsschritt in

Beziehung gesetzt wird (**look-to-buy rate**). Dadurch können die Unternehmen feststellen, ob die Präsentation ihrer Produkte wirkungsvoll ist oder nicht und gegebenenfalls strategische Maßnahmen ergreifen.

### 3.4.3 Recency, Frequency und Monetary Value (RFM)

Die RFM-Methode ermöglicht einen wichtigen Einblick in den Wert eines Kunden oder eines Kundensegments für das jeweilige Unternehmen auf der Grundlage vergangener Transaktionen.

**Recency** beschreibt, wie lange eine Nutzer-Aktion, zum Beispiel Kauf oder Besuch auf der Website, zurückliegt. Die Relevanz dieser Kennzahl liegt in der Tatsache, dass von zurückliegendem Verhalten am besten auf zukünftiges Verhalten geschlossen werden kann. So ist die Wahrscheinlichkeit, dass ein Kunde ein Buch online bestellt größer, wenn der letzte Kauf nicht vor mehreren Monaten, sondern erst kürzlich erfolgt ist.

**Frequency** beschreibt die Häufigkeit, mit der ein Besucher eine bestimmte Website ansteuert oder Käufe in einem Online-Shop tätigt, hat je nach Art des E-Business eine unterschiedliche Bedeutung. In einem typischen B2C-Handlungsumfeld mit geringwertigen Gütern (Blumen, Bücher, Lebensmittel etc.) ist eine hohe Frequenz nahezu gleichbedeutend mit einem höheren Umsatz. Je höher die Besuchsfrequenz, desto höher der Wert des Kunden für das Unternehmen. Dies gilt zwar auch im Handel mit hochwertigen Gebrauchsgütern, die längere Entscheidungsprozesse benötigen, doch indiziert die Frequenz hier etwas anderes. Je häufiger ein potenzieller Kunde die Website, etwa eines Autohändlers besucht, umso kürzer steht er vor der Kaufentscheidung und kann dementsprechend durch gezielte Marketingaktionen angesprochen werden.
In einer kontinuierlichen Kunden-Lieferantenbeziehung wie im B2B-Handel weist ein Wandel im Frequenzmuster von Besuchen oder Bestellungen oft auf eine Änderung des Verhaltens oder der Kaufdisposition hin, zum Beispiel auf eine steigende Unzufriedenheit mit dem Lieferanten.

Die dritte Dimension der Kundenbewertung ist sein finanzieller Wert für das Unternehmen (**Monetary Value**), der sich nach der tatsächlich erzielten Rendite bemisst. Eine „Daumenregel“ besagt, dass mit 10% der Online Kunden 90% des Gewinns erzielt werden. Mit Hilfe des Monetary Value können diese 10% identifiziert, gezielt angesprochen und betreut werden. Streuverluste von Werbekampagnen und Kundenbindungsprogrammen können so vermieden werden. Dies macht den Monetary Value zu einer der wichtigsten Kennzahlen. [vgl. NetGenesis, S.36-38]

Ausgehend von der Analyse des Umsatzverlaufs eines Kunden werden Recency, Frequency und der Monetary Value auf einer Punkteskala bewertet und zu einer Gesamtzahl für den Kundenwert aggregiert. Mit steigender Punktzahl des Kundenwerts kann somit der Erfolg des Kundenbeziehungsmanagements gemessen werden. Vgl Abbildung **3-8**

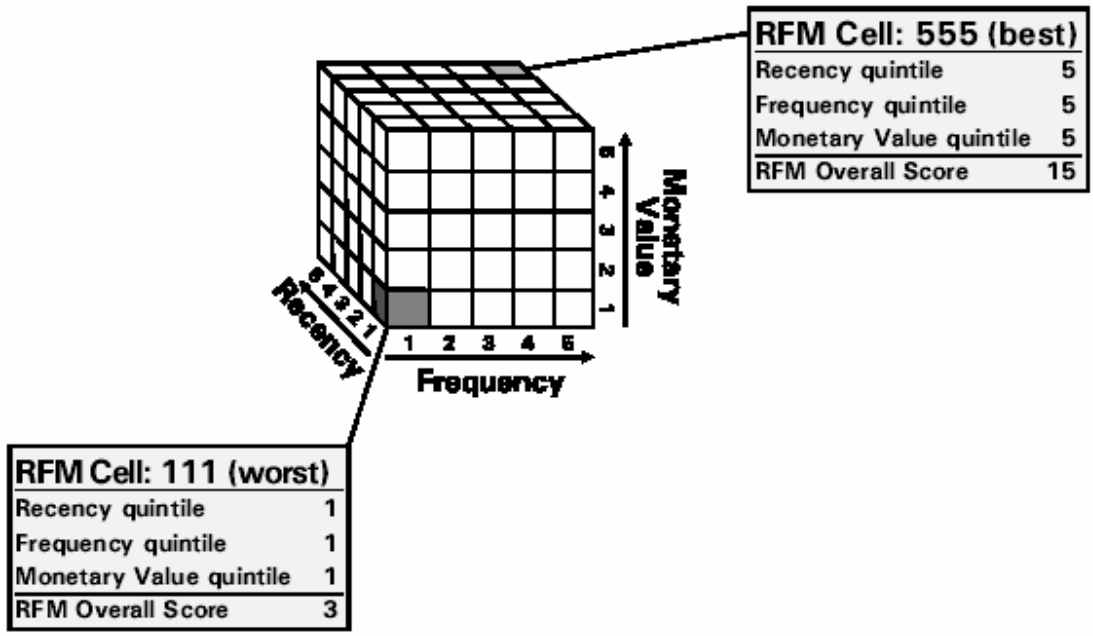



Abbildung **3-8**: Visualisierung der RFM-Methode [NetGenesis, S. 38]

In diesem drei-dimensionalen Würfel werden die Kunden in 125 Segmente eingeteilt. Durch die Betrachtung einzelner Kundensegmente, können gezielte Marketing-Maßnahmen durchgeführt werden, um den Kundenwert, je nach Position innerhalb des Würfels, zu erhöhen.

Wichtig ist jedoch, dass diese Methode immer vergangenheitsbezogen ist, d.h. ein Kunde dessen Potenzial durch eine Erbschaft plötzlich gestiegen oder durch eine Scheidung gesunken ist, kann nicht sofort identifiziert werden, sondern wird erst im nach hinein durch seine historischen Daten "entdeckt". Deshalb sollte dieses Verfahren über einen längeren Zeitraum hinweg durchgeführt werden. In diesem Fall lassen sich nämlich auch Kundenwanderungen zwischen den Zellen feststellen und somit lässt sich der Erfolg der Marketingmaßnahme beurteilen. Ein weiterer Vorteil liegt darin, dass sich diese Methode sehr leicht auf einer Website anwenden lässt.

### 3.5 Kennzahlen zur Analyse des Kundenverhaltens einer Website

Neuere Kennzahlen versuchen stärker, die Interaktivität zur Evaluierung von Websites zu erfassen und stellen somit die unmittelbare Finanzwirksamkeit des Werbeerfolgs in den Mittelpunkt. Dabei wird auf folgende Kennzahlen fokussiert:

**Stickiness**

Mit dieser Kennzahl möchte man die Attraktivität von Websites bewerten, also wie lange und wie häufig sich der Nutzer mit gewissen Inhalten einer Website beschäftigt hat. Sie berechnet sich aus dem Produkt von Besuchsfrequenz, Besuchsdauer und Reichweite.

**Slipperiness**

Im Unterschied zu Stickiness erfasst die Kennzahl die Fähigkeit einer Website zur schnellen Kundeninteraktion mit absichtlich kurzer Verweildauer. Slipperiness ist das Gegenteil von Stickiness und berechnet sich aus denselben Basiswerten und wird als Ausdruck benutzt, wenn die Einzelfaktoren klein sind. Während eine gewisse

Slipperiness gut ist bei solchen Inhalten, bei denen der Nutzer bei zu langem Verweilen auf einer Seite den Besuch abbrechen würde, ist Stickiness in Situationen wünschenswert, in denen lange Verweildauern und hohe Besuchshäufigkeiten (sowie gute Erreichbarkeit) von Vorteil ist.

**Focus**

Diese Kennzahl beschreibt die Anzahl der Sites innerhalb eines bestimmten (thematischen) Bereichs einer Website, die ein Nutzer im Verhältnis zur Anzahl aller Pages des Gesamtangebots besucht. Bei umfangreichen Informationen ist ein hoher Focus etwa in Produktkatalogen positiv. In Sektionen hingegen, die von Besuchern schnell durchlaufen werden sollen, ist es eher hinderlich, eine Vielzahl von Pages aufzurufen. Bei einem Verkaufsvorgang sollte der Focus also eng sein und die Stickiness gering, was nichts anderes bedeutet, dass der User schnell zu seinem Ziel gefunden hat.

**Velocitiy**

Diese Kennzahl misst die durchschnittliche Zeitdauer, die ein Nutzer benötigt, um von der Registrierung auf einer Seite über mehrere Stufen bis zum Kaufabschluss zu gelangen. Die Zeitspanne von der ersten Begegnung mit einem Angebot bis zur letztendlichen Kaufentscheidung bemisst die Effizienz des Angebots. Mit dieser Maßzahl können alternative Klickpfade bis zum Kauf erprobt und miteinander verglichen werden.

**Seducible Moments**

Diese Kennzahl beschreibt Randbedingungen, unter denen ein Kunde besonders empfänglich für ein Angebot ist. Ein bekanntes Beispiel dafür ist der Ein-Klick-Kauf, der Kunden erlaubt, das Objekt der Begierde sofort zu ordern. Seducible Moments beziehen sich nicht nur auf den Produktkauf. Auch wenn es darum geht, Newsletter zu abonnieren oder persönliche Daten zu hinterlassen, sollte die Websitegestaltung Seducible Moments erzeugen. Besonders wichtig sind Seducible Moments dort, wo der Kundentrichter sich beim Übergang in eine neue Phase des Kundenlebens-

zyklus verengt und die Gefahr besteht, dass viele Besucher den Kaufpfad verlassen (vgl. Abbildung **3-6** ). [Strauss, S.117-118]

## 3.6 Kennzahlen zur Erfolgsanalyse von Websites

Durch den Internetauftritt eines Unternehmens kann der Erfolg von Werbe-aufwendungen genau bestimmt werden und den einzelnen Marketing-Maßnahmen entlang des Kundenlebenszyklus zugeordnet werden. Zu den relevanten Kenn-zahlen gehören die Akquisitionskosten, die Konversionskosten und der Netto-Ertrag.

Die Akquisitionskosten pro User ergeben sich aus dem Quotienten der Kosten für eine Kampagne und der dadurch erzielten Click-Throughs. Hat ein Unternehmen 1 Million Banner Ad Impressions für 50.000 Euro eingekauft und damit 5.000 Visits erzielt, also eine Click-Through-Rate von 0,5% generiert, belaufen sich die Akquisitionskosten auf 10 Euro pro akquirierten User.

Ähnlich lassen sich die Konversionskosten ermitteln. Dabei werden die Werbekosten gegen die mit der Kampagne erzielten Verkäufe aufgerechnet. So kann es sein, dass die Akquisitionskosten einer Kampagne relativ hoch sind, die Kosten pro Konversion aber trotzdem niedrig sind, weil die akquirierten Besucher besonders häufig zu zahlenden Kunden werden.

Der Netto-Ertrag ergibt sich aus dem Quotienten von Gesamtkosten und Gesamtresultat einer Werbekampagne. Diese Kennzahl dient der Bewertung komplexerer Prozesse, deren Einzelschritte nicht analysiert werden können. Dies ist beispielsweise der Fall beim Vergleich unterschiedlicher Click-Pfade, die zum gleichen Ziel führen. So lässt sich etwa der Wert von zwei Bannern A und B ermitteln, die jeweils unterschiedliche Click-Through- und Konversionsraten aufweisen.

Bislang konnten nur „einfache“ Metriken wie die Click-Through- oder Weiterleitungsquote mit den getätigten Werbeinvestitionen verglichen werden. Auf Basis der Kombination dieser Kennzahlen sind jetzt komplette Return-on-Investment-Rechnungen möglich. [NetGenesis, S.40-41]

## 3.7 Komplexe Kennzahlen: Personalisierungsindex, Life Time Value, Loyality Value, Freshness Factor

Der **Personalisierungsindex** misst, inwieweit vorhandene Personalisierungselemente (d.h. Informationen, die über den Kunden bekannt sind) auch wirklich auf der Website genutzt werden. Er berechnet sich aus dem Quotienten der Anzahl in der Kundeninteraktion genutzten Personalisierungselemente und der Anzahl erfasster Personalisierungselemente. Ein Personalisierungsindex muss nicht unbedingt auch einen hohen Personalisierungsgrad in der Ansprache bedeuten. Denn wenn nur zwei Personalisierungselemente erfasst und beide genutzt werden, ist der Personalisierungsindex mit 1 zwar optimal, aber die zwei Elemente an sich sind nicht hinreichend für eine personalisierte Ansprache.

Der **Life Time Value** ist im Grunde genommen keine originäre E-Metrik, kann aber durch den Einsatz des Internets verbessert werden. Er basiert auf einer Gewinn/Verlust-Rechnung pro Kunde. Zu Anfang wird der Wert negativ sein, durch die ersten Käufe positiv werden und dann, falls weitere Käufe ausbleiben, wieder absinken. Websites können dazu genutzt werden, die Kosten pro Kunde zu senken z.B. durch Verlagerung von Service-Leistungen ins Internet oder um Wiederholungskäufe zu initiieren und damit den Life Time Value zu steigern.

Der **Loyalty Value** gehört zu den komplexesten E-Metriken. Er gibt an, wie treu die Kunden als Ganzes sind und wie verlässlich man mit ihnen rechnen kann. Er gehört

zu den entscheidenden Faktoren, mit denen man den Wert eines Online-Unternehmens ermittelt.
Eine einheitliche Definition oder gar eine verbindliche Berechnungsformel gibt es nicht, kann es auch gar nicht geben, da Loyalität für jede Website und jedes Unternehmen etwas anderes bedeutet. Wenn ein User einen Email Account bei Yahoo und einen bei Hotmail hat, und beide täglich benutzt, welchen Wert hat der User dann für das Unternehmen hinsichtlich seiner Loyalität? So ergab zum Beispiel eine Studie von Bain & Company (www.Bain.com), dass Amazon einen Kunden erst nach zweieinhalb Jahren Loyalität zu einem profitablen Kunden einstuft.

Der Loyalty Value muss von jedem Unternehmen als multidimensionaler, eindeutiger Index konstruiert werden, der kontinuierlich und zeitnah die Loyalitäts-werte misst. Für ein Webunternehmen, dass mit geringerwertigen Produkten handelt, könnten die wesentlichen Dimensionen für den Index folgende sein: Visit Frequenz, Visit Dauer, Visit Tiefe, Käufe pro Visit, Anzahl gekaufter Produkte pro Visit, Umsatz pro Visit und Profitabilität pro Monat.

Mit dem **Freshness Factor** möchte man beurteilen, wie sich das Ändern von Inhalten der Website auf die Besucherhäufigkeit auswirkt. Lange bestand die Annahme, dass sich eine Website ständig ändern muss, um das Interesse der Besucher zu halten. Während dies für Websites sinnvoll ist, die Nachrichten, Wet-terprognosen, Sportergebnisse etc. enthalten und somit täglich genutzt werden, stellt die fortlaufende Pflege und Aktualisierung von Inhalten einen bedeutenden Kostenfaktor für ein Unternehmen dar. Demzufolge stellt sich hier schnell die Frage nach der Effektivität und dem Return on Investment. Der Freshness Factor hilft bei der Analyse, indem er die durchschnittliche Modifizierungsrate des Inhalts mit der durchschnittlichen Visit-Häufigkeit vergleicht.

Ist der Faktor kleiner 1 sehen die Besucher mehrmals den gleichen Inhalt, weil sie öfter kommen als der Inhalt ausgetauscht wird, ist er größer 1 finden sie jedes Mal

eine neue Website vor. Geht der Wert über 1,5 werden Ressourcen verschwendet, da manche Inhalte gar nicht wahrgenommen werden, da sie schon wieder verändert und aktualisiert wurden, bevor der durchschnittliche Besucher sie gesehen hat. [NetGenesis, S.51-57]

# 4 Fazit

Derzeit werden als hauptsächliche Datenquelle Logfiles zur Erfolgsbeurteilung einer Website verwendet. Die daraus generierten Kennzahlen sind jedoch oft technischer Natur und haben keinen oder nur einen geringen wirtschaftlichen Bezug. Letztendlich kann der Erfolg eines Internetauftritts nicht nur am erzielten „Traffic“ festgemacht werden, sondern muss auch die Intensität der Kundenbeziehung berücksichtigen. Diese E-Metriken der Zukunft können E-Business Managern helfen, die profitabelsten Kunden und Kundensegmente zu identifizieren, Werbeausgaben intelligent zu steuern, sowie den Kundenlebenszyklus und Return-on-Investment zu analysieren und zu optimieren.

Bei der Erfassung und Sammlung von Daten sind Restriktionen der Internet-Technologie und gesetzliche Rahmenbedingungen zu beachten. Die Hauptprobleme sind dabei die Nutzeridentifikation sowie die eindeutige Abgrenzung und vollständige Erfassung von Nutzungsvorgängen.

Ein einheitliches, allgemein anerkanntes Kennzahlensystem zur Be- und Auswertung der Datenerhebung existiert derzeit nicht. Zusammenfassend ist festzustellen, dass das Controlling des Internetauftritts ein unabdingbarer Bestandteil der Webstrategie eines Unternehmens sein sollte, um Kosten und Nutzenaspekte einzelner Maßnahmen und des gesamten Web-Auftritts beurteilen zu können.

## Literaturverzeichnis:

Exody E-business Intelligent GmbH. E-commerce Controlling http://www.olap-competence-center.de/.../$File/ e-commerce%20controlling%20(exody).pdf, 07.07.2003

Gomory S, Hoch R, Lee J, Podlaseck M, Schonberg E. E-Commerce Intelligence: Measuring, Analyzing and Reporting on Merchandising Effectiveness of Online Stores http://www.research.ibm.com/iac/papers/eabs3.pdf, 19.05.2003

Hippner H, Merzenich M, Wilde KD. Handbuch Web Mining im Marketing,1. Auflage, Braunschweig: Friedr. Vieweg & Sohn Verlagsgesellschaft mbH 2002

Müller A, Thienen L. Controlling von e-Business-Systemen mit der Balanced Scorecard. http://www.competence-site.de/site.de/controlling.nsf/4E521C9FDE1E3B9AC1256BA3006B2A88/$File/e-business_balancedscorecard.pdf, 10. Juni 2003

NetGenesis Corp. E-Metrics, Business Metrics for the new economy http://maya.cs.depaul.edu/~classes/ect584/papers/e-metrics.pdf, 19.05.2003

Ringlein T, Hahn P, Mitterer B, Schneppe C, Ederer F. E-business-Anwendungen und die Notwendigkeit im Umdenken des Unternehmens im Controlling. Controller Magazin 3/2002: 283-290

Schwickert AC, Wendt P. Controlling-Kennzahlen für Web Sites. http:/www/wi.uni-giessen.de/gi/dl/showfile/Schwickert/1168/Apap_WI_2000_08.pdf, 08. Juni 2003

Strauss R, Schoder D. e-Reality. Frankfurt am Main: F.A.Z.-Institut für Management-,Markt-und Medieninformationen GmbH 2002

# Data Mining Standards am Beispiel von PMML

Markus Prechtel

04/08/03

# Inhaltsverzeichnis


**I Data Mining** **2**

**1 Einleitung** **2**

**2 Allgemeine Definitionen** **3**

**3 Der Mining Prozess nach CRISP-DM** **5**

**II PMML** **8**

**4 PMML als Standard im Data Mining** **8**
4.1 Eine kurze Motivation für Standards im Data Mining allgemein 8
4.2 Anforderungen an Standards . . . . . 9
4.3 Der Standard PMML . . . . . 10

**5 Der Aufbau von PMML** **11**
5.1 Allgemein . . . . . 11
5.2 Syntax von PMML am Beispiel eines Entscheidungsbaumes . . 13

**6 Die Verbreitung von PMML** **17**

# Teil I
# Data Mining

## 1 Einleitung

Daten werden heutzutage in immer umfangreicherem Maße erhoben. Dies geschieht auf sehr unterschiedliche Art und Weise und in sehr unterschiedlichen Gebieten, wie z.B. in der Qualitätssicherung oder im Kundenmanagement: Marktforscher tragen Informationen über Kundenverhalten zusammen, medizinische Untersuchungsergebnisse werden gesammelt, Produktionsprozesse werden mittels Sensoren protokolliert usw.
Die große Herausforderung für Unternehmen liegt nun darin, aus den Rohdaten interessante und nützliche Informationen zu generieren. Ein nützliches Werkzeug dazu stellt das "Schürfen" in den Daten, das sogenannte Data Mining dar. Darunter versteht man, nicht offensichtliche Zusammenhänge, Muster und Strukturen in den Daten zu erkennen. [7]
Die daraus gewonnenen Erkenntnisse werden dann u.a. dazu verwendet, die Kundenansprache individueller zu gestalten, das Kundenverhalten vorherzusagen oder um die Qualität der Produkte zu verbessern bzw. um Schwachstellen aufzudecken.
Data Mining-Verfahren verbinden Methoden aus den Bereichen Statistik, künstliche Intelligenz, Datenbankanwendungen und Visualisierung, um den Prozess des Auffindens verborgener und interessanter Informationen in großen Datenbeständen umfassend zu unterstützen. Sie erkennen Trends und Muster in Daten, indem sie automatisch eine Vielzahl von möglichen Abhängigkeiten generieren und so wichtige Zusammenhänge innerhalb der Daten extrahieren können.

# 2 Allgemeine Definitionen

**Data Mining**
'Ein Prozess , um interessante neue Muster, Korrelationen und Trends in großen Datenbeständen zu entdecken, um damit entscheidungsunterstützende Prognose-Modelle zu erstellen.' [2]

**Knowledge Discovery in Databases (KDD)**
Im Gegensatz zum Data Mining, das lediglich die Suche nach auffälligen Mustern darstellt, umfasst KDD den gesamten Findungsprozess (von der Frageformulierung bis hin zur Ergebnisinterpretation). Data Mining ist somit ein Teilschritt des KDD.
Knowledge Discovery beschreibt automatisierte Verfahren, mit denen (unter Anwendung von Data Mining Methoden) Regelmäßigkeiten in Mengen von Datensätzen aufgefunden und in eine für den Anwender verständliche Form gebracht werden. [4]

**Data Warehouse**
Prinzipiell ist das Data Warehouse ein Datenspeicher, welcher die zentrale Sammelstelle für Unternehmensdaten darstellt und diese sinnvoll strukturiert und organisiert. Es bietet jedem Mitglied eines Unternehmens die Möglichkeit, sich Information gemäß den individuellen Bedürfnissen zu besorgen, was insbesondere auch für das Management der Firma von großem Interesse ist.
Im Unterschied zu konventionellen Datenbanken speichert das Data Warehouse Informationen in der Regel ewig, d.h. ein Löschen der Daten ist normalerweise nicht möglich. [4]

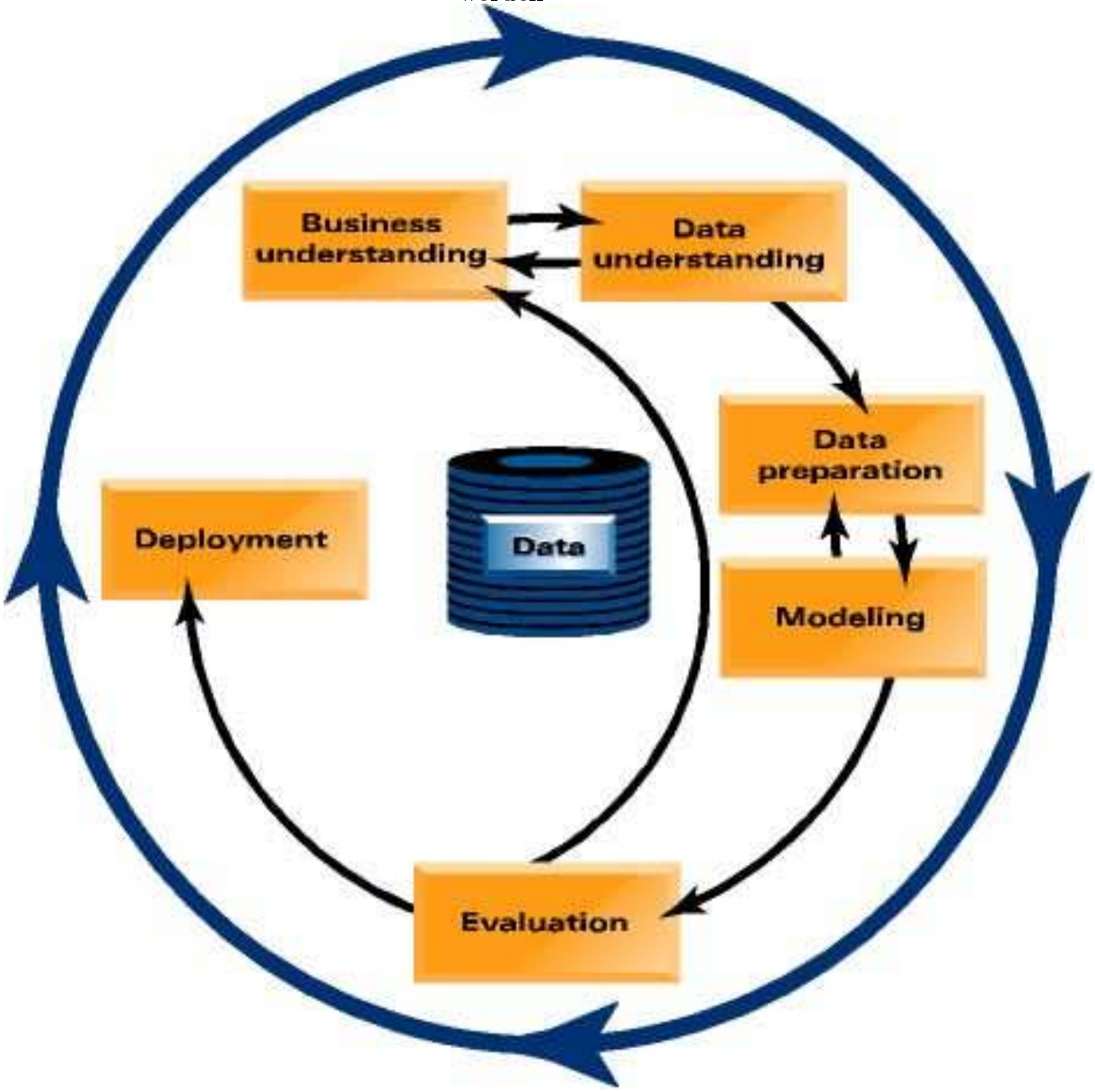
**On-Line Analytical Processing (OLAP)**
In großen Data Warehouse Umgebungen stoßen einfache Abfrage-Werkzeuge für den Datenzugriff schnell an die Grenzen ihrer Leistungsfähigkeit, und der Performance-Gewinn wird zum zentralen Thema. Ein weiteres Problem liegt im eingeschränkten Funktionsumfang von SQL. Komplexe analytische Abfragen, wie beispielsweise "Zeige mir alle regionalen Verkäufe als Prozentzahl in NRW von Februar bis März 1999, und vergleiche es mit dem gleichen Zeitraum im letzten und vorletzten Jahr", sind mit Standard-SQL nur schwer durchzuführen und dann nicht sehr benutzerfreundlich.
Beim OLAP handelt es sich letztendlich um eine Technik, die es dem jeweiligen Anwender erlaubt, sich mittels eines interaktiven Zugriffs in eine Vielzahl von Sichten und Darstellungsweisen auf Basisdaten einen schnellen Einblick zu verschaffen.
OLAP ist die konzeptionelle Basis für Lösungen zur Unterstützung der dy-

namischen Analyse von Unternehmensdaten. Es geht hierbei nicht darum, die klassischen Datenbanken zu ersetzen, sondern darum, die auf operativer Ebene etablierten klassischen Datenbanken mit analytischen Funktionen zu ergänzen.
Im Prozess der Entscheidungsfindung können die Daten nicht nur von einer, sondern von mehreren Perspektiven aus analysiert werden. Durch das multidimensionale Modell können die Unternehmensdaten dann von verschiedenen Positionen betrachtet und sofort graphisch umgesetzt werden.
Somit ist OLAP ein Begriff, der oft in Verbindung mit Wissensfindung auftritt und sich mit Fragen wie 'Welches Produkt wurde wann wo am besten verkauft?' beschäftigt. (Im Gegensatz zum Data Mining, was sich auf den Grund solcher Ereignisse bzw. Tatsachen bezieht.) [9]

**Data Mart**
Data Marts können als Subsysteme eines Data Warehouse angesehen werden, welche lediglich auf einen oder nur wenige Betriebsbereiche beschränkt sind. Ein Data Mart kann auch alleine für Data Mining oder OLAP Zwecke benützt werden. [4]

# 3 Der Mining Prozess nach CRISP-DM

CRISP-DM ist ein von ein paar Unternehmen (Terradata Copenhagen, Denmark (ein Unternehmensbereich von NCR); Integral Solutions Limited (ISL), United Kingdom (ein Teil von SPSS); DaimlerChrysler Aktiengesellschaft, Germany; OHRA Verzekering en Bankk Groep B.V., Netherlands) zusammengeschlossenes Konsortium, das es sich zur Aufgabe gemacht hat, den Data Mining-Prozess zu standardisieren. Das Ergebnis der Entwicklung soll hier nur kurz veranschaulicht werden:



Darstellung des Data Mining Prozesses nach CRISP-DM [8]

Der Lebenszyklus eines Data Mining Projektes besteht aus sechs Phasen. Die zeitliche Abfolge der Phasen ist hierbei nicht streng festgelegt, ein vor- und zurückspringen zwischen verschiedenen Phasen wird immer benötigt. Es hängt vom Ergebnis jeder einzelnen Phase ab, welche Phase als nächstes durchlaufen werden soll, wobei man im Hinterkopf behalten sollte, dass nicht immer das gesamte Projekt mit einem einzigen Software Werkzeug durchgeführt wird, sondern man nach Möglichkeit für jede Phase das jeweils beste einsetzt. Die dabei auftretenden Schnittstellen wiederum solten hierbei natürlich durch entsprechende Standards definiert werden. Im folgenden sollen die einzelnen Phasen kurz erläutert werden:

**Business Understanding**
Diese Anfangsphase konzentriert sich auf die Formulierung des eigentlichen Projektziels aus der Perspektive des Anwenders und konvertiert dieses Wissen dann in eine Data Mining-Problemdefinition und einen vorläufigen Plan, um diese Zielsetzungen zu erreichen.

**Data Understanding**
Hier wird mit der Datenauswahl begonnen, gefolgt von Aktivitäten mit dem Ziel, die Daten besser kennenzulernen und Probleme der Datenqualität oder interessante Teilmengen der Daten zu entdecken.

**Data Preparation**
Die nächste Stufe beinhaltet Aktivitäten zur Generierung der endgültigen Analysedaten. Darin enthalten ist die Auswahl von Daten, sowohl von Variablen als auch von Datensätzen, wie auch Transformationen und das Entfernen von "Datenschmutz".

**Modelling**
Auf dieser Stufe werden verschiedene Verfahren ausgewählt und angewandt, sowie die zugehörigen Parameter festgesetzt. Typischerweise gibt es mehrere Techniken für denselben Data Mining Problemtyp. Manche haben spezifische Voraussetzungen an die Form der Daten, weshalb man an dieser Stelle oftmals zur Data Preparation Stufe zurückgehen muss.

**Evaluation**
Zu Beginn dieser Phase hat man bereits ein Modell aufgestellt, das aus Sicht der Datenanalyse hohe Qualität hat. Bevor man nun jedoch zur endgültigen Auswertung des Modells in der Deployment-Phase kommt, ist es wichtig das

Modell zu bewerten und die einzelnen Schritte, die gemacht wurden, um es zu konstruieren, genau zu prüfen. Dies geschieht, um sicher zu gehen, dass es genau den geforderten Voraussetzungen gerecht wird.

[8] und [6]

# Teil II
# PMML

## 4 PMML als Standard im Data Mining

### 4.1 Eine kurze Motivation für Standards im Data Mining allgemein

Ein Data Mining Prozess wird heutzutage meistens als kleiner Teil eines größeren Prozesses durchgeführt. Da sowohl die Daten, die dem Data Mining als Input dienen, oftmals in sehr unterschiedlicher Form gegeben sind, als auch der Output des Data Mining Prozesses in verschiedensten Formen benötigt wird, entstehen oftmals Probleme bei der Verwendung verschiedener Software Tools. Es ist aber von großem Vorteil, für jedes Problem, das beste Werkzeug zu verwenden. Die jeweils für ein Problem besten Werkzeuge stammen jedoch nicht immer von einem Anbieter und da es sich in der Regel um kommerzielle Software handelt, bleibt ein problemloser Datenaustausch oftmals ein Wunschtraum.
Die Data Mining Modelle und die statistischen Modelle, die von kommerzieller Data Mining Software erzeugt werden, sind somit oftmals nur Komponenten in anderen Systemen, wie beispielsweise im Customer Relationship Management, in der Unternehmensplanung, im Risikomanagement usw. Man kann die Integration, das Aktualisieren und die Instandhaltung solcher Data Mining Software und der Systeme enorm vereinfachen, indem man Data Mining Standards einsetzt.
Schon eingeführte und aufstrebende Standards richten sich an einige Aspekte des Data Mining, z.B.:

- *Modelle:*
  Um Data Mining und statistische Daten zu repräsentieren, was den Austausch von Modellen vereinfacht.

- *Attribute:*
  Um die Säuberung, die Transformation und das Sammeln von Attributen, die als Input in die Modelle verwendet werden, darzustellen. Diese Standards richten sich somit an die Teilbereiche "Modelling" (Modellierung) und "Data Preparation" (Datenaufbereitung) aus dem letzten Kapitel.

- *Interfaces und APIs:*
  Um eine Verbindung zu anderen Sprachen und Systemen zu schaffen.

Z.B. enthält der SQL Multimedia and Applications Packages Standard einen Teil "SQL/MM Part 6: Data Mining", in welchem eine Schnittstelle geschaffen wird, um Data Mining Anwendungen direkt von SQL-fähigen relationalen Datenbanken zu erreichen. [3]

- *Einstellungen:*
  Um die internen Parameter abzubilden, die benötigt werden um die Modelle aufzubauen und zu benutzen.

- *Prozess:*
  Um die Modelle herzustellen und mit ihnen zu arbeiten, d.h. dass der gesamte Data Mining-Prozess, wie er z.B. von CRISP-DM beschrieben wird, standardisiert wird.

[12]

## 4.2 Anforderungen an Standards

Der Hauptgrund, weshalb heutzutage so viele Standards existieren, ist die Tatsache, dass Data Mining in so vielen verschiedenen Gebieten und in Kombination mit so vielen verschiedenen Systemen und Diensten verwendet wird, von denen viele ihre eigenen, oft inkompatiblen Standards benötigen. [12]
Es ergeben sich somit einige grundlegende Überlegungen für die weitere Entwicklung von Standars:

- Der Austausch von Daten zwischen unterschiedlichen Programmen bzw. Software Tools muss problemlos ablaufen können.

- Die Data Mining Modelle sollten ebenfalls standardisiert unter verschiedenen Anwendungen weitergegeben werden können.

- Standards sollten es ermöglichen, die Modelle in anderen Softwareprodukten integriert anzuwenden.

Mittlerweile führen jedoch zwei Aufgaben die Data Mining Standard-Agenda an: sich auf einen einheitlichen Standard für die Datenbereinigung, -transformation und -aufbereitung zu einigen (PMML Version 2.0 repräsentiert einen ersten Schritt in diese Richtung) und sich auf eine einheitliche Menge von Web Diensten zu einigen, um mit Fern- oder verteilten Daten zu arbeiten. Letzteres ist eine Bemühung, deren Realisierung eben erst begonnen hat. [12]

## 4.3 Der Standard PMML

Wörtlich übernommen lautet die Begriffsdefinition von PMML:
*'The Predictive Model Markup Language (PMML) is an XML mark up language to describe statistical and data mining models.'* [5]
("Die Textauszeichnungssprache für Vorhersagemodelle PMML ist eine XML-basierte Textauszeichnungssprache um statistische Modelle und Data Mining Modelle zu beschreiben.")
PMML wurde von der Data Mining Group (http://www.dmg.org), einer 1998 gegründeten Vereinigung einiger Unternehmen (darunter IBM, Microsoft, Oracle), entwickelt, um Data Mining und statistische Modelle darzustellen und zu beschreiben. PMML beinhaltet auch die Beschreibung der benötigten Operationen für die Datenbereinigung und -aufbereitung.
Ein Hauptziel war es, verschiedenen Anwendungen und OLAP Werkzeugen Zugriff zu Modellen zu verschaffen, die aus verschiedenen Quellen erhalten werden, ohne sich dabei um die Unterschiede dieser Quellen Gedanken machen zu müssen. [10]
PMML versucht somit, eine ausreichende Infrastruktur bereitzustellen, sodass eine Anwendung in der Lage ist, ein Modell zu konstruieren (der sog. PMML Produzent) und eine andere Anwendung nur durch das Lesen der PMML XML Datei dieses benutzt (der sog. PMML Konsument). [12]
Dazu dokumentiert PMML die Eingaben in Data Mining Modelle, die Transformationen, die zuvor für die Datenaufbereitung benutzt wurden und die Parameter, die die Modelle selbst beschreiben. [5]
Mit Hilfe von PMML werden Modelle erstellt, von denen Vorhersagemodelle ('Predictive Models') abgeleitet werden können. In den letzten Jahren wurden viele Vorhersagemodelle entwickelt, die nun durch PMML einfach und plattformunabhängig miteinander ausgetauscht und verglichen werden können. [10]
Somit erfüllt PMML zumindest einen Teil der Forderungen, die für die gesamte Data Mining Welt von immenser Wichtigkeit sind.

# 5 Der Aufbau von PMML

## 5.1 Allgemein

PMML besteht aus den folgenden Komponenten:

- **Datenbeschreibungsverzeichnis** (Data dictionary)
  Definiert die Input-Attribute der Modelle und spezifiziert die verschiedenen Datentypen und deren Bereiche und bezieht sich auf die Phase "Business Understanding" im Prozessmodell nach CRISP-DM.
  Es beginnt immer mit `<DataDictionary ... >` (wobei "..." durch weitere Attribute wie z.B. `numberOfFields = "5"` ersetzt wird) und endet mit `</DataDictionary>` . Dazwischen kommen die Definitionen der einzelnen Datenfelder incl. ihrer Definitionsbereiche (die von verschiedensten Typen sein können):
  ```
  <DataField name="Temperatur" optype="continuous"/>
  ```
  Ein Datenfeld, dessen Definitionsbereich direkt im Anschluss angegeben wird, muss hierbei mit `</DataField>` abgeschlossen werden:
  ```
  <DataField name="Wind" optype="categorical">
  <Value value="true"/>
  <Value value="false"/>
  </DataField>
  ```

- **Mining Schema**
  Es existiert genau ein Mining Schema in jedem Modell, welches die Attribute und deren Funktion im Modell auflistet. Diese Attribute sind eine Teilmenge der Attribute im Datenbeschreibungsverzeichnis. Das Schema enthält spezifische Informationen für ein bestimmtes Modell, während das Datenbeschreibungsverzeichnis Daten-Definitionen beinhaltet, die für alle Modelle gleich sind. Es gibt zusätzlich noch den Gebrauchstyp ('usage type') eines Attributes an, der aktiv (ein Input des Modells), vorhergesagt(ein Output des Modells) oder ergänzend (enthält nur beschreibende Zusatzinformationen und wird vom Modell ignoriert) sein kann. Das Mining Schema beginnt mit `<MiningSchema>` und endet jeweils mit `</MiningSchema>` . Analog zum Data Dictionary stehen dazwischen sogenannte Mining Fields, die von der Art `<MiningField name="Temperatur"/>` sind.
  Es bezieht sich auf die Phasen "Data Understanding" und "Data Preparation" im CRISP-DM Modell.

- **Transformationsverzeichnis** (Transformation dictionary)
  Kommt ebenfalls in der CRISP-DM-Phase "Data Preparation" zum

Tragen und kann eine der folgenden Transformationen beinhalten und ist zumindest bei manchen Modellen optional:

- Normalisierung (normalization): Unter Normalisierung versteht man die Abbildung von Werten in einen vorgegebenen Bereich (gewöhnlicherweise das Intervall $[0, 1]$ ). Sie wird beim Arbeiten mit neuronalen Netzen verwendet, in Regressionsmodellen gebraucht man jedoch oft auch ähnliche Verfahren.
- Diskretisierung (discretization): Darunter versteht man die Abbildung von Intervallen auf Zeichenketten, i.e. "Worte". Z.B. kann der Wertebereich des Gewinns einer Firma in die zwei Intervalle $(-\infty, 0)$ und $[0, +\infty)$ unterteilt und diese beiden Intervalle können dann auf {"gut", "schlecht"} abgebildet werden.
- Wertzuordnung (value mapping): Jede Zahl und jede Zeichenkette kann einer anderen Zahl oder einer anderen Zeichenkette zugeordnet werden, indem man in einer Tabelle einfach alle Paare angibt. Z.B. kann "weiblich" auf "w" und "männlich" auf "m" abgebildet werden.
- Verdichtung (aggregation): Werte werden gruppiert und/oder aufaddiert, ähnlich wie beim Berechnen von arithmetischen Mitteln oder der Gruppierung bei SQL mit "GROUP BY".

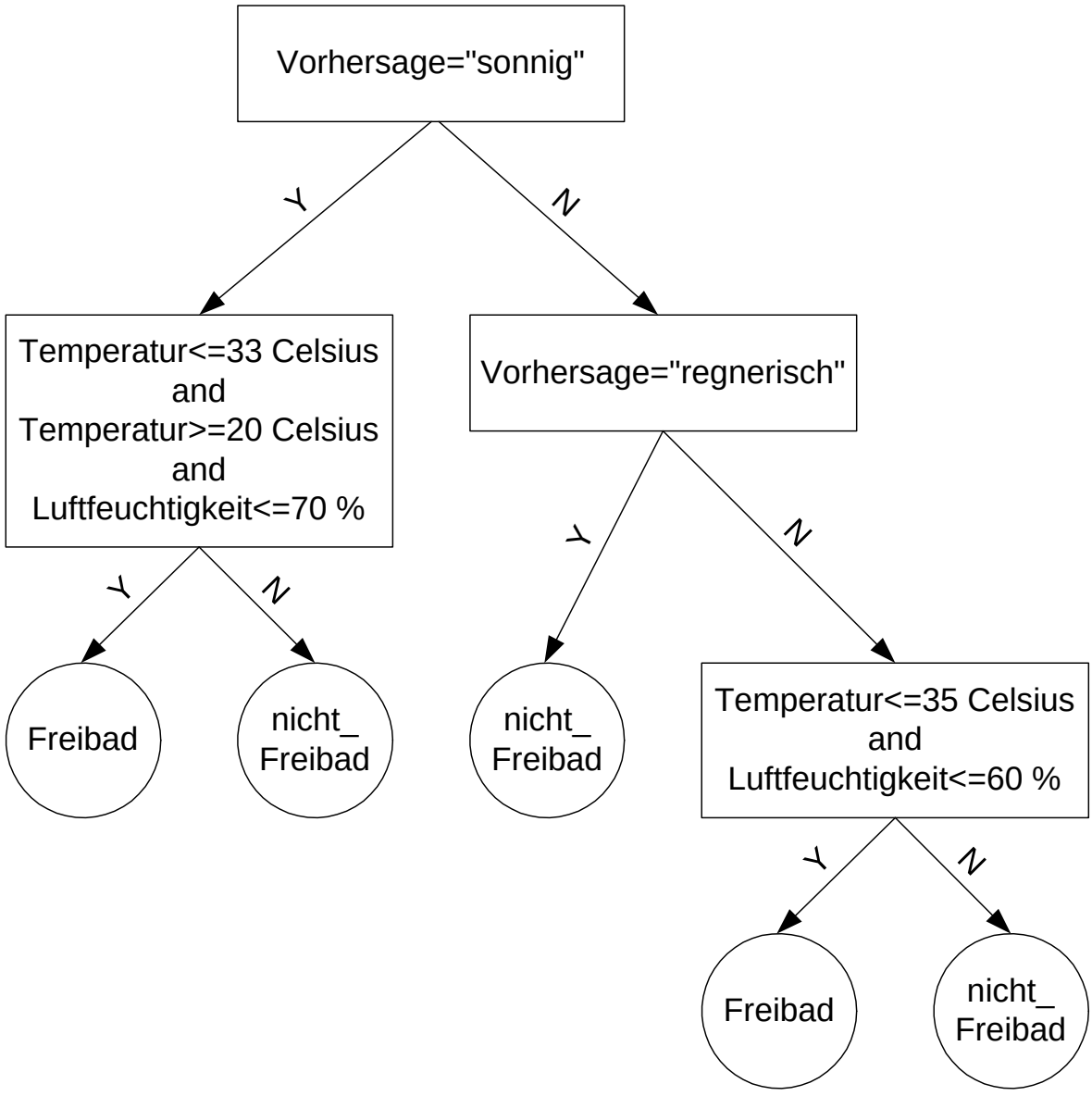
[5]

- **Modellstatistiken** (Model statistics): Beinhaltet univariate Statistiken über die Attribute im Modell und ist ebenfalls in manchen Modellen optional. Dieser und der nächste Punkt beziehen sich auf die Stufe "Modelling" im CRISP-DM Modell.
- **Modelle**: Enthält die Modellparameter, die u.a. den Modelltyp spezifizieren. PMML 2.0 enthält Regressionsmodelle, Clustermodelle, Bäume, neuronale Netze, Bayes'sche Modelle, Assoziationsregeln und Sequenzmodelle. Die genaue Syntax kann man leicht anhand des Beispiels im nächsten Abschnitt erkennen.

[12]

## 5.2 Syntax von PMML am Beispiel eines Entscheidungsbaumes

Der Aufbau eines Entscheidungsbaumes mit PMML soll anhand eines Beispiels erläutert werden. Zuerst eine Skizze des zu erzeugenden Binärbaumes:




Beispiel eines Entscheidungsbaumes

Die Syntax, mit der man einen solchen Baum erzeugen kann, ist nicht besonders kompliziert und auch mit wenigen XML Kenntnissen leicht einsehbar und verständlich:

```
1<?xml version="1.0"?> <PMML version="1.1">
2   <Header description="Ein kleiner Entscheidungsbaum"/>
3   <DataDictionary numberOfFields="5">
4      <DataField name="Temperatur" optype="continuous"/>
5      <DataField name="Luftfeuchtigkeit" optype="continuous"/>
6      <DataField name="Wind" optype="categorical">
7         <Value value="true"/>
8         <Value value="false"/>
9     </DataField>
10     <DataField name="Vorhersage" optype="categorical">
11        <Value value="sonnig"/>
12        <Value value="bedeckt"/>
13        <Value value="regnerisch"/>
14     </DataField>
15     <DataField name="whatIdo" optype="categorical">
16        <Value value="Freibad"/>
17        <Value value="nicht_Freibad"/>
18     </DataField>
19  </DataDictionary>
20  <TreeModel modelName="Schwimmen_gehen">
21     <MiningSchema>
22        <MiningField name="Temperatur"/>
23        <MiningField name="Luftfeuchtigkeit"/>
24        <MiningField name="Wind"/>
25        <MiningField name="Vorhersage"/>
26        <MiningField name="whatIdo" usageType="predicted"/>
27     </MiningSchema>
28     <Node score="Freibad">
29        <Predicate field="Vorhersage" operator="equal" value="sonnig"/>
30        <Node score="Freibad">
31           <CompoundPredicate booleanOperator="and">
32              <Predicate field="Temperatur" operator="lessThan" value="33 Celsius"/>
33              <Predicate field="Temperatur" operator="greaterThan" value="20 Celsius"/>
34              <Predicate field="Luftfeuchtigkeit" operator="lessThan" value="70 % "/>
35           </CompoundPredicate>
36           <Node score="Freibad"> <True/> </Node>
```

```
37          <Node score="nicht_Freibad"> <True/> </Node>
38       </Node>
39       <Node score="Freibad">
40          <Predicate field="Vorhersage" operator="equal" value="regnerisch"/>
41          <Node score="nicht_Freibad"> <True/> </Node>
42          <Node score="Freibad">
43             <Predicate field="Wind" operator="equal" value="true"/>
44             <Node score="nicht_Freibad"> <True/> </Node>
45             <Node score="Freibad">
46                <CompoundPredicate booleanOperator="and">
47                   <Predicate field="Temperatur" operator="lessThan" value="35 Celsius"/>
48                   <Predicate field="Luftfeuchtigkeit" operator="lessThan" value="60 % "/>
49                </CompoundPredicate>
50                <Node score="Freibad"> <True/> </Node>
51                <Node score="nicht_Freibad"> <True/> </Node>
52             </Node>
53          </Node>
54       </Node>
55    </Node>
56 </TreeModel>
57</PMML>
```

Jede Definition eines Entscheidungsbaumes beginnt mit `<TreeModel ...>` . Der Name des Modells kann im optionalen Attribut "modelName" angegeben werden, im Beispiel heit das Modell Schwimmen_gehen. Jedes Modell enthält ein Mining Schema, optional Statistiken, im Beispiel nicht vorhanden, und die einzelnen Knoten `<Node score="Freibad">` des Entscheidungsbaumes. Die Knoten ihrerseits bestehen aus einem Prädikat
`<Predicate field="Vorhersage"operator="equal"value="sonnig"/>` , das alle Objekte, die in dem Knoten selbst, oder einem seiner Unterknoten enthalten sind, erfüllen müssen, sowie weiteren Knoten, sofern der Knoten selbst kein Blattknoten ist. Bei den Prädikaten handelt es sich entweder um Terme der Form `<Feldname> <operator> <Wert>` oder um eine Verknüpfung
`<CompoundPredicate booleanOperator="and">` solcher Terme durch die boolschen Operatoren "and", "or" oder "xor". Blattknoten werden gekennzeichnet, indem sie anstatt eines Prädikats einen Wahrheitswert enthalten.
Als Operatoren innerhalb der Prädikate sind "equal" ($=$), "notEqual" ($\neq$), "lessThan" ($<$), "lessOrEqual" ($\leq$), "greaterThan" ($>$) und "greaterOrEqual" ($\geq$) erlaubt. Das vorgeschriebene Attribut "score" innerhalb des Elementes "Node" bezeichnet den Wert, den das Modell an diesem Knoten für das Zielfeld vorhersagt. Bei Blattknoten stellt dieser Wert gleichzeitig die

Klassenbezeichnung dar.
Das Mining Schema des Beispiels enthält die Felder "Temperatur", "Luftfeuchtigkeit", "Wind", "Vorhersage" und "whatIdo". Das Feld "whatIdo" ist das Zielfeld, welches durch das Modell vorhergesagt werden soll. [1]

# 6 Die Verbreitung von PMML

Momentan ist PMML zwar schon im Vergleich zu anderen Data Mining Standards sehr weit verbreitet, absolut betrachtet jedoch noch nicht. Die Ausbreitung dieses von allen Beteiligten sehr begrüßten Standards nimmt aber sehr schnell zu.
Mittlerweile ist er von vielen führenden Unternehmen in ihre Produkte integriert. Dazu gehören Oracle, Microsoft (OLE DB) und IBM (intelligent miner). Es gibt auch einige Softwarelösungen, mit denen eine Schnittstelle zu Java, C++ und CORBA (z.B. mit Xelopes von Prudsys [11]) aufgebaut wird.
Auch SQL bietet mit der ISO/IEC 13249 (SQL Multimedia and Applications Packages Standard) einen Teil (SQL/MM Part 6), mit dem Data Mining Konzepte in ein Data Warehouse integriert (Stichwort "embedded Data Mining") werden können und der voraussichtlich im Herbst diesen Jahres veröffentlicht wird. Hierbei wird auch PMML unterstützt. [3]
Es ist somit anzunehmen, dass der positive Trend der letzten Jahre ("hin zu PMML") auch in Zukunft fortgesetzt wird. Dies wird garantiert auch durch die Version PMML 2.0, die im März 2003 von der Data Mining Group auf den Markt gebracht wurde, unterstützt. Und da auch schon auf der Homepage von DMG ( http://www.dmg.org ) verkündet wird:
'Work on PMML Version 3.0 has begun'
ist die Frage, welcher Standard im Data Mining Bereich in den kommenden Jahren am weitesten verbreitet sein wird, vermutlich größtenteils geklärt.

# Literatur


[1] Björn Stadler. Austauschformate für Data Mining und OLAP. May 2001.

[2] Pasquale Borelli. Information Mining für CRM basierend auf der MicroStrategy 7 Plattform. 2001.

[3] Christoph Lingenfelder. SQL/MM Part 6: Data Mining, Jul 2003.

[4] Daniela Jakwerth, Robert Kratky, Wilhelm Reither. Datenschutz vs. Data Mining.

[5] Data Mining Group.

[6] Dr. Lars Schmidt-Thieme. Web Mining. Nov 2000.

[7] Lehrstuhl für ABWL und Wirtschaftsinformatik Katholische Universität Eichstätt-Ingolstadt. Data Mining - Mehr Gewinn mit Kundendaten.

[8] http://www.crisp-dm.org.

[9] http://www.mkonetzny.de.

[10] http://www.oasis-open.org.

[11] http://www.xelopes.de.

[12] R.L.Grossman, M.F.Hornick, G.Meyer. Data Mining standard initiatives. *COMMUNICATIONS OF THE ACM*, (Vol.45, No.8), Aug 2002.

Seminar:
Internet-Dienste

Prof. Dr. Franz Schweiggert

Thema:

# E-Marketing

WWW und E-Mail
im Rahmen der
Kommunikationspolitik

Von Daniel Nießner

## Inhaltsverzeichnis

1. Einleitung 3

2. Grundlagen des E-Marketing 3
- 2.1 Begriffsdefinition 3
- 2.2 Vor- und Nachteile des E-Marketing 4
- 2.3 Voraussetzungen für effektives E-Marketing 6
- 2.4 Push- und Pull-Werbung 6

3. Formen des E-Marketing 7
- 3.1 Webseiten 7
  - 3.1.1 Basiserfolgsfaktoren 7
  - 3.1.2 Webseiten Gestaltungsfaktoren 8
- 3.2 Online-Werbung 8
- 3.3 E-Mail-Marketing 10
  - 3.3.1 Vorteile des E-Mail-Marketing 10
  - 3.3.2 Permission-Marketing 11
    - 3.3.2.1 Single Opt-in 12
    - 3.3.2.2 Confirmed Opt-in 12
    - 3.3.2.3 Double Opt-in 12
  - 3.3.3 Spamming 13
    - 3.3.1 Rechtslage 13
    - 3.3.2 Schutz vor Spam-Mails 14

4. Ausblick 14

5. Literaturverzeichnis 15

# 1. Einleitung

Im Jahr 2002 nutzten über 50 Prozent der Deutschen das Internet, 60 Prozent der Internetnutzer sind fast tagtäglich online. Gerade deshalb ist es für Unternehmen mittlerweile unerlässlich die klassischen Marketinginstrumente durch E-Marketing zu ergänzen. Auch wenn nach einer Studie der Nielsen Media Research GmbH im Jahr 2002 auf das E-Marketing nur 1,5 Prozent der gesamten Werbeausgaben entfielen, wird das Internet als Werbemedium weiter an Bedeutung gewinnen. Nach der Studie wurden im Jahr 2002 in Deutschland 255,9 Millionen Euro für Werbung im Internet ausgegeben. Das entspricht einem Plus von 20,9 Prozent gegenüber dem Vorjahr. Die Prognos AG geht von einer Verdopplung der Werbeausgaben bis ins Jahr 2007 aus.

Diese Seminararbeit beschränkt sich auf die populärsten Internet-Dienste, die für das E-Marketing die größte Bedeutung haben: World Wide Web (WWW) und E-Mail. Es wird nur die Kommunikationspolitik als E-Marketinginstrument betrachtet und die Einsatzmöglichkeiten von WWW und E-Mail aufgezeigt. Der Schwerpunkt liegt dabei auf der in Deutschland meistgenutzten E-Marketing-Form, dem E-Mail-Marketing. Laut einer Studie der DoubleClick GmbH Deutschland im Jahr 2002 unter Marketing-Verantwortlichen nutzten 63 Prozent der Unternehmen E-Mail-Marketing.

# 2. Grundlagen des E-Marketing

## 2.1 Begriffsdefinition

Unter E-Marketing versteht man die interaktive, kommerzielle Kommunikation mit Individuen oder Massen mittels vernetzter Informationssysteme. E-Marketing bedeutet die Übertragung des klassischen Marketing auf elektronische Medien, also hauptsächlich E-Mail und WWW. Das Ziel des E-Marketing ist wie beim klassischen Marketing, neue Kunden zu gewinnen und sie an das Unternehmen zu binden. Das Verhältnis von E-Marketing zum klassischen Marketing muss jedes Unternehmen für sich erarbeiten, umsetzen und laufend aktualisieren. Welche Bedeutung E-Marketing für ein Unternehmen hat hängt letztlich von der Branche, dem Produkt und den Bedürfnissen der Zielgruppen ab. [BMWA]

## 2.2 Vor- und Nachteile des E-Marketing

**Vorteile** [Warschburger: 11; BMWA]

***Direkte Feed-back-Möglichkeiten***
Kunden können zum Beispiel Bestellungen sofort aufgeben, an Diskussionsforen teilnehmen oder Nachfragen zu E-Mail-Informationen abschicken.

***Informationen über den Kunden***
Mit Hilfe von Abrufstatistiken für Webseiten oder auch durch den Einsatz von Cookies, lassen sich Informationen über die Kunden sammeln und für die weitere Planung auswerten.

***Schnelligkeit***
Kundeninformationen, wie zum Beispiel Informationen per E-Mail oder ein neuer Online-Katalog, können in kurzer Zeit veröffentlicht werden. Das Internet-Angebot kann schnell an die Bedürfnisse der Zielgruppe angepasst werden.

***Grenzenlose Reichweite***
Internet und E-Mail haben sich mittlerweile als Kommunikationsform etabliert mit der potenzielle Kunden regional, national und international angesprochen werden können. Beide werden sowohl für die Geschäftsabwicklung zwischen Unternehmen und Verbraucher (B2C) als auch für die Geschäftsabwicklung zwischen Unternehmen (B2B) genutzt.

***Ständige Verfügbarkeit***
Das Internetangebot eines Unternehmens steht den Internetnutzern rund um die Uhr öffentlich zur Verfügung.

***Kostenersparnis***
E-Marketing verursacht weniger Kosten als das klassische Marketing. Während im klassischen Marketing ein großer Teil der Kosten durch den Druck von Katalogen, Schaltung von Anzeigen oder Werbespots oder auch Porto für Postsendungen verursacht wird, entfallen diese Kosten beim E-Marketing. Zwar entstehen beim E-Marketing auch Kosten wie zum Beispiel für die Bearbeitung der Online-Texte und für die Beantwortung von E-Mails, aber es ergibt sich insgesamt ein Preisvorteil gegenüber der klassischen Werbung.

***Anschaulichkeit***
Kundeninformationen können durch multimediale Elemente, wie zum Beispiel Animationen, Video oder Ton, anschaulicher und ansprechender präsentiert werden.

**Nachteile** [Warschburger: 12]

***Schwellenängste***
Bestimmte Zielgruppen sind gegenüber den neuen Technologien abgeneigt. Vor allem ältere und nicht-berufstätige Personen haben oft nicht die nötigen Kenntnisse im Umgang mit PC und Software und trauen sich die Nutzung eines PCs nicht zu.

***Erreichbarkeit***
Bestimmte Zielgruppen sind noch nicht erreichbar. Beispielsweise nutzt nach einer Forsa-Umfrage nur circa ein Viertel der über Fünfzigjährigen das Internet.

***Sicherheitsbedenken***
Viele Nutzer haben Bedenken bezüglich Datenschutz, Zahlungsabwicklung oder Rechtssicherheit im Internet.

***Wettbewerbsdruck***
Dem Internet-Nutzer ist es möglich sich genauestens über ein Unternehmen zu informieren und sich ein Bild davon zu machen. Mit dieser verstärkten Transparenz erhöht sich auch der Wettbewerbsdruck.

***Preisdruck***
Der Nutzer hat die Möglichkeit schnell und bequem Preise zu vergleichen. Diese unmittelbare Vergleichbarkeit, die das Internet dem Nutzer bietet, führt zu Preisdruck.

***Kundenbindung***
Eine langfristige Kundenbindung wird zunehmend schwieriger, da durch die verkürzten und transparenten Kundenwege via Mausklick eine hohe Kundenwechselgefahr besteht.

## 2.3 Voraussetzungen für effektives E-Marketing

***Public Relations***
Die potenziellen Kunden müssen über das Online-Angebot informiert werden. Internet- und E-Mail-Adressen sollten in allen Offline-Werbemaßnahmen des Unternehmens erscheinen und in alle Geschäftspapiere aufgenommen werden. Falls dem Unternehmen die E-Mail-Adressen seiner Kunden bekannt sind, sollte es diese per E-Mail informieren.

***Qualifikation***
Die Mitarbeiter sollten fundierte Kenntnisse sowohl im klassischen Marketing als auch im E-Marketing und in den Anwenderprogrammen haben, notfalls müssen sie entsprechend geschult werden.

***Technik***
Durch das Versenden von Massen-Mails werden Mail-Server erheblich belastet. Daher sollte bei der Auswahl des Mail-Servers die Zahl der Mails, die Größe der Dateien und die Häufigkeit des Mail-Versands berücksichtigt werden.

***Zielgruppenorientierung***
Die Webseite des Unternehmens sollte nicht nur attraktiv gestaltet und intuitiv bedienbar sein, sondern auch auf die Zielgruppe ausgerichtet sein und diese auch wirklich erreichen.

***Präsentation des Unternehmens***
Das Unternehmen sollte sich im Internet seinen potenziellen Kunden als vertrauenswürdiger Geschäfts- und Dienstleistungspartner präsentieren um sie an sich zu binden.

## 2.4 Push- und Pull-Werbung

Die Besonderheit der Kommunikationspolitik als E-Marketinginstrument liegt in der Pull-Kommunikation, bei der der Nutzer selbst entscheidet, welches Informationsangebot im Internet er wahrnimmt und wann er dies tut. Dadurch verspricht man sich eine höhere Akzeptanz von Werbung, da die Nutzer nicht durch unerwünschte Werbung belästigt werden. Im Gegensatz dazu ist die Werbung in den traditionellen Medien durch einen Push-Charakter gekennzeichnet. Die Zuschauer, Zuhörer oder Leser werden unaufgefordert mit Werbung konfrontiert. Allerdings hält das Push-Prinzip auch im Internet Einzug. In zunehmendem Maße wird der Nutzer durch unverlangte

Werbeinformationen belästigt. Dazu zählen beispielsweise Pop-Ups, Interstitials und Spam-Mails. [Warschburger: 239 ff.]

# 3. Formen des E-Marketing

## 3.1 Webseiten

Eine Webseite bietet einem Unternehmen die Möglichkeit sich und seine Produkte oder Leistungen einfach und weltweit zu präsentieren. Auf seiner Webseite kann das Unternehmen die Zielgruppen mit produkt- und unternehmensspezifischen Informationen versorgen. In der Regel besteht der Internet-Auftritt eines Unternehmens aus einer Homepage und dahinter liegenden Seiten. Die für den Erfolg einer Webseite entscheidenden Faktoren lassen sich in Basiserfolgsfaktoren und Webseitengestaltungsfaktoren unterteilen. [Warschburger: 229 ff.]

### 3.1.1 Basiserfolgsfaktoren

***Aktualität***
Eine Webseite sollte laufend aktualisiert werden, da aktuelle Informationen das Interesse des Nutzers wecken und somit Anreize für einen wiederholten Besuch der Webseite schaffen.

***„Verwicklung des Nutzers“ (Involvement)***
Der Internet-Auftritt sollte den Nutzer dazu animieren, länger auf der Webseite zu verweilen, den Angeboten zu folgen und darin zu „blättern“.

***Designqualität***
Es muss das richtige Verhältnis von Inhalt und grafischen Elementen gefunden werden, da einerseits das Internet kein typisches Lesemedium ist und deshalb die Texte eher kurz sein sollten und andererseits die Datenübertragungsrate nicht zu stark durch viele multimediale Elemente wie Animationen oder Grafiken belastet werden sollte.

***Technische Erreichbarkeit***
Die Wartezeit für den Webseiten-Aufbau sollte für den Nutzer akzeptabel sein. Studien zufolge liegt die maximale Wartezeit zwischen 10 und 30 Sekunden.

### 3.1.2 Webseiten-Gestaltungsfaktoren

***Value Added Services***
Sie sollen den Nutzer auf die Webseite locken und an sie binden. Beispiele für Value Added Services sind Online-Spiele, Online-Gewinnspiele, Free-Services wie kostenloser SMS Versand, Chat- und Diskussionsforen oder Links zu anderen interessanten Seiten.

***Zugänglichkeit***
Dem Nutzer muss eine klare und strukturierte Seitennavigation geboten werden, damit er sich auf den Webseiten zurechtfindet. Die Zugänglichkeit kann durch zur Verfügung gestellte Suchfunktionen erhöht werden, da der Nutzer so einfacher zu den von ihm gewünschten Ergebnissen kommt.

***Möglichkeit zum direkten Kontakt***
Neben Ansprechpartnern und Telefonnummern sollten auch Webformulare zur direkten Kommunikation oder direkte Bestellmöglichkeiten angeboten werden.

## 3.2 Online-Werbung

Damit eine Webseite erfolgreich wird, muss sie zunächst einmal den alten und neuen Kunden bekannt gemacht werden. Dies sollte sowohl durch Werbung in den traditionellen Medien als auch durch Online-Werbung erfolgen. Die Internet Adresse sollte in TV-Spots, Prospekten, Anzeigen in Zeitungen und Zeitschriften, Katalogen und allen Geschäftsunterlagen eingebunden werden. Im Internet wird die Werbung meist auf stark frequentierten Webseiten anderer Unternehmen platziert. Dafür eignen sich besonders die Webseiten von Online-Zeitungen und Online-Zeitschriften, Suchmaschinen und Portale. [BMWA]

**Online-Werbeformen** [Fritz: 143 ff.; VDZ: 12 ff.]

***Textlinks***
Die einfachste Form der Vermarktung einer Webseite sind Textlinks auf einer fremden Webseite. Nach dem Anklicken des Links gelangt der Internet-Nutzer auf die beworbene Webseite.

***Banner***
Die bekannteste Form der Online-Werbung sind Banner. Ein Banner ist grundsätzlich eine mit einem Link versehene rechteckige Grafik, von der aus der Betrachter durch

Anklicken direkt auf die beworbene Webseite gelangt. Obwohl die Klickraten sinken, bilden Banner einen elementaren Bestandteil jeder Online-Werbekampagne, unter anderem deshalb, weil sich für Banner schon früh Standardformate etabliert haben.

***Bannertausch***

Hierbei werden Banner von Kooperationspartnern oder Mitgliedern in Bannertausch-Netzwerken auf der eigenen Webseite eingeblendet. Dafür erhält man in einem bestimmten Tauschverhältnis Einblendungen auf anderen Seiten. Der Bannertausch stellt eine kostenlose Alternative zur bezahlten Schaltung von Bannern dar.

***Buttons***

Buttons sind eine Sonderform der Banner. Durch ihre zurückhaltende Gestaltung, sie enthalten nur den Namen eines Unternehmens oder eines Produkts, wirken sie weniger störend und lassen sich besser in das Layout der Webseite integrieren.

***Scroll Ads oder Sticky Ads***

Scroll Ads oder Sticky Ads sind Banner, die mit dem Nutzer „wandern“, wenn er auf einer Webseite nach unten scrollt. So bleiben sie immer im Blickfeld des Nutzers.

***Pop-Ups***

Ein Pop-Up-Advertisement legt sich beim Aufruf einer Internet-Seite in einem eigenen kleinen Fenster über deren Inhalt und kann erst durch einen Mausklick wieder geschlossen werden.

***Interstitials***

Beim Aufruf einer bestimmten Webseite, erhält der Nutzer statt der gewünschten Zielseite zuerst eine bildschirmfüllende Werbung zu sehen. Erst nach einer gewissen Zeit oder einem Klick gelangt der Nutzer zur gewünschten Webseite. Man bezeichnet dies als „Unterbrecherwerbung“.

***Rectangle***

Der Rectangle ist eine horizontale Online-Anzeige in einem relativ großen Format, die in den unmittelbaren Contentbereich integriert ist. Er eignet sich besonders als Branding Instrument.

***Micro-Sites***

Innerhalb eines Banners wird eine kleine, vollfunktionsfähige Webseite eingeblendet. Oft enthalten Micro-Sites nicht nur Produktinformationen, sondern auch Bestellmöglichkeiten und werden deshalb auch als Transactive Banner oder Interactive Banner bezeichnet.

***Keyword-Advertising***
Viele Suchmaschinen ermöglichen so genanntes Keyword-Advertising, bei dem die Bannereinblendung vom Suchbegriff abhängt, den der Nutzer eingibt.

***Content-Sponsoring***
Beim Content-Sponsoring wird beispielsweise das Logo eines Sponsors in die Webseite des Gesponsorten eingebunden und dient als Link zu der Webseite des Sponsors.

***Platzierung in Suchmaschinen***
Die Chance einer Webseite, von der Zielgruppe gefunden zu werden, hängt in hohem Maße von der Platzierung in Suchmaschinen ab. Diese geben dem Internet-Nutzer Hinweise auf relevante Webseiten, nachdem dieser bestimmte Suchbegriffe in die Suchmaschine eingegeben hat.

## 3.3 E-Mail-Marketing

E-Mail ist der meistgenutzte Dienst im Internet. Zahlreiche Untersuchungen belegen, dass der größte Teil der Internet-Nutzer auch E-Mails empfängt und versendet. Man kann E-Mail-Marketing als die gezielte Versendung von Informationen und Werbebotschaften per E-Mail definieren. Für Unternehmen stellt es einen schnellen und kostengünstigen Weg dar um Kundenbeziehungen auf- und auszubauen, da es sich ideal für Direkt- und Dialogmarketing mit Interessenten und (potenziellen) Kunden eignet. E-Mail-Marketing kann auf vielfältige Weise eingesetzt werden: in Form von Newslettern und Diskussionslisten, als Erinnerungsservice oder für die unterschiedlichsten Informationsangebote. Beim Einsatz von E-Mail-Marketing muss grundsätzlich zwischen Permission-Marketing, bei dem der Kunde zustimmt E-Mails zu empfangen, und Spamming, bei dem die E-Mails ohne Erlaubnis des Empfängers verschickt werden, unterschieden werden.

### 3.3.1 Vorteile von E-Mail-Marketing

***Günstige Kosten***
Die Distributionskosten für E-Mails sind deutlich günstiger als das Porto für Postsendungen. Außerdem fallen keine Ausgaben für Papier oder Druck an.

***Schnelligkeit***
E-Mails gehen meist innerhalb weniger Sekunden nach dem Versand beim Empfänger ein.

***Rücklaufquote***
Das Antworten auf eine E-Mail ist wesentlich einfacher und bequemer als auf eine Post- oder Faxsendung. Deshalb sind die Rücklaufqouten auch deutlich höher.

***Messbarkeit***
Alle Reaktionen auf eine E-Mail, zum Beispiel wer wann wie oft auf welche Links in der E-Mail geklickt hat, lassen sich elektronisch erfassen und automatisch exakt auswerten. Mit den Ergebnissen können nachfolgende E-Mail-Marketing-Aktionen optimiert werden.

***Potential zum One-to-One-Marketing***
Die Inhalte von E-Mails lassen sich individuell auf das Profil des jeweiligen E-Mail-Empfängers anpassen. Es kann ein inhaltlich und zeitlich individueller persönlich auf den E-Mail-Empfänger zugeschnittener E-Mail-Dialog geführt werden.

***Interaktivität***
E-Mails im HTML- oder Flash-Format können interaktive Elemente wie zum Beispiel Webformulare enthalten und lassen sich damit aktiver und eindringlicher gestalten.

***Multimediale Elemente***
E-Mails im HTML- oder Flash-Format können multimediale Elemente, wie zum Beispiel Bilder, Animationen oder Musik, enthalten um eine höhere Aufmerksamkeit zu erzielen.

### 3.3.2 Permission-Marketing

Unerwünschte E-Mails werden oft nicht nur ungelesen gelöscht, sie können auch gegen rechtliche Vorschriften verstoßen. Dieses Problem kann durch das sogenannte Permission-Marketing, das auf dem Einverständnis des Empfängers eine E-Mail zu empfangen beruht, gelöst werden. Letztendlich kann das Permission-Marketing zu einer höheren Kundenbindung führen, da es einem Anbieter ermöglicht personalisierte Angebote zu erstellen, die vom Empfänger zumindest nicht unerwünscht sind. Allerdings muss der Empfänger auf jeden Fall die Möglichkeit haben, die Erlaubnis zum Empfang von E-Mails jederzeit wieder zurücknehmen zu können. Bei der Anmeldung zu einem E-Mail-Verteiler oder einem E-Mail-Newsletter gibt es mehrere

Registrierungsverfahren. Allen gemeinsam ist, dass der Interessent auf der Webseite des Anbieters seine E-Mail-Adresse und eventuell weitere Angaben über ein Webformular an den Anbieter schickt.

Für das anschließende Registrierungsverfahren gibt es drei Varianten: das „Single Opt-in“, das „Confirmed Opt-in“ und das „Double Opt-in“. [Aschoff: 30 ff.]

#### 3.3.2.1 Single Opt-in (auch Opt-in)

Die vom Webformular übermittelten Daten werden ausgewertet und die E-Mail-Adresse des Empfängers in die Verteilerliste aufgenommen. Ein Problem besteht allerdings darin, dass Dritte E-Mail-Adressen anderer Personen anmelden und diese dann unerwünschte E-Mails erhalten.

#### 3.3.2.2 Confirmed Opt-in

Der E-Mail-Empfänger erhält eine Bestätigungs-Mail, die einerseits die Anmeldung bestätigt und andererseits einen Link für die umgehende Abmeldung enthält. Falls die E-Mail-Adresse von einem Dritten eingetragen wurde, kann sich der Empfänger über den Link direkt wieder aus der Verteilerliste austragen. Es kann davon ausgegangen werden, dass der Empfänger die Anmeldung akzeptiert hat, wenn er nicht innerhalb einer bestimmten Zeitspanne, beispielsweise eine Woche, auf die Bestätigungs-Mail reagiert. Auch beim Confirmed Opt-in-Verfahren besteht die Möglichkeit dass ein Dritter E-Mail-Adressen anderer Personen anmeldet.

#### 3.3.2.3 Double Opt-in

Auch beim Double Opt-in erhält der Empfänger eine Bestätigungs-Mail, in der er allerdings aufgefordert wird die E-Mail als Gegenbestätigung wieder zurückzuschicken. Erst wenn der Absender die Gegenbestätigung erhält wird die E-Mail-Adresse des Empfängers in die Verteilerliste aufgenommen. Somit wird sichergestellt, das die E-Mail-Adresse tatsächlich existiert und dass nur Personen die E-Mail erhalten, die es auch ausdrücklich wünschen. Aus Marketing-Sicht hat das Double Opt-in-Verfahren aber einen entscheidenden Nachteil: Ein Teil der E-Mail-Empfänger wird aus Bequemlichkeit, Unkenntnis oder Unvermögen die Gegenbestätigungs-Mail nicht zurücksenden und geht als Abonnent verloren. Genau deshalb ist das Double Opt-in-Verfahren wenig verbreitet, obwohl es höchste Rechtssicherheit bietet.

### 3.3.3 Spamming

Spam ist eigentlich der Markenname für das gepökelte und gepresste Dosenfleisch der US-Firma Hormel Foods Corp. und die Abkürzung für „spiced pork and ham“. Aufgrund seiner umstrittenen Qualität wurde Spam von den britischen Komikern von Monty Python’s aufgegriffen. In einem Sketch der Monty Python’s ging es um ein Restaurant, in dem jedes Gericht Spam enthält. Egal was ein Gast bestellt, es ist immer Spam dabei, obwohl es der Gast nicht ausstehen kann. Anscheinend waren unter den ersten Internet-Nutzern viele Fans von Monty Python’s, denn als in den Diskussionsforen die ersten sich massenhaft wiederholenden unerwünschten Werbebotschaften auftauchten, setzte sich schnell der Begriff Spam dafür durch. Als dann begonnen wurde, die Werbebotschaften per E-Mail massenhaft zu verschicken, wurde der Begriff Spam auf unerwünschte E-Mails ausgeweitet. Nach der Definition der EU-Kommission versteht man heute unter Spam unverlangt zugestellte E-Mails. Synonym werden auch die Begriffe „Junk Mail“, „Bulk Mail“ oder UCE („Unsolicited Commercial E-Mail“) verwendet. Die meisten der Spam-Mails sind kommerziell und werden wegen der geringen Kosten in großen Massen verschickt. [Aschoff: 137]

#### 3.3.3.1 Rechtslage

In Deutschland ist das Versenden von Werbung über E-Mail ohne Zustimmung des Empfängers rechtswidrig. Bereits in der einmalig unverlangten Zusendung von Werbung per E-Mail an Privatpersonen wird eine Verletzung des allgemeinen Persönlichkeitsrechts (LG Berlin, 16 O 201/98) gesehen. Auch gegenüber Gewerbetreibenden sieht die Rechtssprechung die Zusendung von Spam als rechtswidrig an. Das Europäische Parlament hat einer Richtlinie zugestimmt nach der E-Mail-Werbung von Unternehmen nur mit Erlaubnis der Empfänger versandt werden darf. Die Richtlinie muss in den EU-Staaten bis Ende Oktober 2003 umgesetzt werden. Anders sieht es in den USA aus. Ein erheblicher Teil der Spam-Angriffe wird von US-amerikanischen Auftraggebern vorgenommen, da es dort bisher kein einheitliches Anti-Spam-Bundesrecht gibt. Allerdings haben bereits einige Bundesstaaten Gesetze erlassen, die dieses Thema behandeln. [eco: 3]

Abgesehen von den rechtlichen Folgen von Spam-Mails sinkt auch die Akzeptanz der E-Mail-Empfänger gegenüber Marketing- und Werbe-Mails mit der Anzahl der unerwünschten Spam-Mails, und die Akzeptanz von Werbe-Mails ist schließlich entscheidend für den Erfolg von E-Mail-Marketing.

#### 3.3.3.2 Schutz vor Spam-Mails

Um sich gegen Spam-Mails zu schützen, gibt es für einen E-Mail-Empfänger mehrere Möglichkeiten. Auf jeden Fall sinnvoll ist das Einrichten eines alternativen E-Mail-Accounts für bestimmte Aktivitäten im Internet, wie zum Beispiel die Teilnahme an Gewinnspielen oder beim Schreiben in Diskussionslisten. Ausserdem bietet sich der Einsatz von Filterprogrammen an. Eine weitaus umstrittenere Möglichkeit ist die Eintragung in die so genannte E-Mail-Robinsonliste, die den E-Mail-Empfänger vor unerwünschten E-Mails schützen soll. Ob dies aber die Versender von Spam-Mails davon abhält, den E-Mail-Empfänger mit der unerwünschten Werbung zu verschonen, ist äußerst fraglich. Manchmal ist es auch möglich die Spam-Mails zurück zu verfolgen und gegen den Absender vorzugehen. [eco: 5 ff.]

Auch Unternehmen sollten bestrebt sein, nicht als Absender von Spam-Mails eingestuft zu werden, da dies dem Image des Unternehmens erheblich schaden könnte. Vor dem Kauf oder Anmieten von E-Mail-Adressen von einem Adressbroker sollte das Unternehmen prüfen beziehungsweise sich vom Adressbroker schriftlich zusichern lassen, dass das Einverständnis der E-Mail-Empfänger vorhanden ist. Bei einem Tausch von E-Mail-Adressen mit einem anderen Unternehmen muss beachtet werden, dass das Einverständnis der E-Mail-Empfänger nur für das Unternehmen gilt, das die E-Mail-Adressen ursprünglich gewonnen hat. Ein Austausch ist nur dann sinnvoll, wenn die E-Mail-Empfänger beider Unternehmen einem Adresstausch zugestimmt haben. Vor dem Versand von E-Mails sollte das Unternehmen seine E-Mail-Verteilerliste mit der E-Mail-Robinsonliste oder gegebenenfalls einer eigenen Sperrliste abgleichen, um so zu verhindern, dass Personen E-Mails erhalten, die es nicht wünschen. Unternehmen die befürchten, dass Dritte die E-Mail-Adressen anderer Personen anmelden, sollten für die Anmeldung zu ihrer E-Mail-Verteilerliste das „Double Opt-in“-Verfahren statt dem „Single Opt-in“-Verfahren verwenden, da dies sicherstellt, dass nur Personen E-Mails erhalten, die es auch ausdrücklich wünschen.

## 4. Ausblick

Die Individualisierung des Informationsangebots im Internet wird weiter vorangetrieben werden, um die Kundenbindung weiter zu verbessern. Auch der Trend im E-Mail-Marketing geht vom Massen-Marketing, bei dem jeder E-Mail-Empfänger die gleiche E-Mail erhält, hin zum One-to-one-Marketing, bei dem jeder E-Mail-Empfänger eine inhaltlich und zeitlich individuelle, persönlich auf ihn zugeschnittene E-Mail erhält. Erst mit dem Medium E-Mail ist One-to-one-Marketing im großen Stil möglich. [Aschoff: 193 ff.]

# 5. Literaturverzeichnis

Aschoff, Martin: Professionelles Direkt- und Dialogmarketing per E-Mail, 2002 München

Bundesministerium für Wirtschaft und Arbeit (BMWA): Unternehmer – E-Business http://www.bmwi.de/Navigation/Unternehmer/E-Business /marketing.html, 12.07.2003

Conrady / Jaspersen / Pepels (Hrsg.): Online Marketing Instrumente, 2002 Neuwied und Kriftel

Fritz, Wolfgang: Internet-Marketing und Electronic Commerce, 2. Auflage, 2001 Wiesbaden

Sonntag, Dr. Ralph: Herausforderungen beim Online-Marketing unter besonderer Berücksichtigung des Permission Marketing, 12/2002, http://www.competence-site.de/marketing.nsf/80B9EC59CD AD6F9AC1256CB4002FE6B4/$File/online_marketing.pdf, 19.07.2003

VDZ Verband Deutscher Zeitschriftenverleger e.V. (Hrsg.): Online Werbung – Fakten und Perspektiven, 2002 Berlin, http://www.vdz.de/owfb/index.html, 10.07.2003

VDZ Verband Deutscher Zeitschriftenverleger e.V. (Hrsg.): Branchendaten New Media 2003, 2003 Berlin, http://www.vdz.de/pages/static/181.aspx, 12.07.2003

Verband der deutschen Internetwirtschaft e.V.: eco – Electronic Commerce Forum: Spam, 2003 Köln, http://www.eco.de/servlet/PB/menu /1015951/index.html

Walter, Volker: Die Zukunft des Online-Marketing, 2. Auflage, 2000 München

Warschburger, Volker / Jost, Christian: Nachhaltig erfolgreiches E-Marketing, 1. Auflage, 2001 Braunschweig / Wiesbaden

# E-COMMERCE

Seminararbeit im Vertiefungsfach Informatik

Studiengang Wirtschaftswissenschaften

Sommersemester 2003

Universität Ulm

Bearbeiterin: Sylvia Reichardt

Matrikelnummer: 427088

Betreuer: Dr. Melzer

Abgabe: 30. Juli 2003

Inhalt:

1. Einleitung

2. Wirtschaftliche Aspekte
2.1 Was ist e-commerce
2.2 Was macht online-shopping attraktiv
2.2.1 für die Konsumenten
2.2.2 für die Anbieter
2.3 Erfolgsfaktoren
2.3.1 Allgemeine Erfolgsfaktoren
2.3.2 Unternehmensspezifische Erfolgsfaktoren
2.3.2.1 Art der angebotenen Waren und Dienstleistungen
2.3.2.2 Versandhandel
2.3.2.3 Dienstleistungen
2.3.2.4 Portale und Marktplätze
2.3.2.5 Fachgeschäfte

3. Vertrauen im e-commerce

4. Probleme im e-commerce
4.1 Soziale Probleme
4.2 Technisch/Rechtliche Probleme
4.2.1 Regelungen des BGB
4.2.2 Elektronisches Geschäftsverkehr Gesetz
4.2.3 Signaturgesetz
4.3 Zahlungsverfahren und Systeme

5. Zusammenfassung, Aktueller Stand und Ausblick

## 1. Einleitung

Die in den letzten Jahren stattgefundene Weiterentwicklung in den Bereichen Computer und Internet ist auch an der Geschäftswelt nicht vorbeigegangen. Es entstanden neue Möglichkeiten der Vermarktung, insbesondere über das WorldWideWeb. Nach einer anfänglichen Euphorie mussten jedoch viele Unternehmen erkennen, dass der e-commerce bei weitem nicht für jedes Produkt, jede Branche und jeden Kunden gleich geeignet ist. Zum Beispiel sind Versandhandelsunternehmen, Banken und Reisevermittler erfolgreicher als andere Branchen und es werden vor allem jüngere Kunden durch den e-commerce angesprochen.
Im weiteren Verlauf werde ich deshalb auf Unterschiede zwischen den einzelnen Produkten, erfolgsbildende Faktoren, sowie die derzeit noch bestehenden Probleme im Zusammenhang mit dem e-commerce eingehen.

# 1. Wirtschaftliche Aspekte

## 1.1. Was ist e-commerce?

In aktuellen Lehrbüchern existiert eine Vielzahl verschiedener Begriffsdefintionen für den e-commerce. Dieses Problem entsteht vor allem aus der Unklarheit darüber, welche Geschäftstätigkeiten im e-commerce enthalten sind und welche nicht dazu zählen. Die Definitionen reichen von sehr engen , wie z.B. dem blossen „Online-Verkauf von Produkten und Dienstleistungen“[1] über die „Anwendung der Informationstechnologie zur Unterstützung von Geschäftsprozessen und dem Austauch von Gütern und Dienstleistungen“[2] bis hin zu weiteren Ausdrücken wie der „Unterstützung von Geschäftstransaktionen, Geschäftsprozessen sowie der Beziehungen zu sämtlichen internen und externen Partnern eines Unternehmens durch Informations- und Kommunikationstechnologie“[3].
Ein anderer Ansatzpunkt liegt darin, den e-commerce nicht als neuen Markt zu definieren, sondern lediglich als neuen Distributionskanal über den Produkte und Dienstleistungen abgesetzt werden. Die Besonderheit dabei ist, dass Angebot sowie die Bestellung über das Internet abgewickelt werden.
Der wichtigste Unterschied zwischen dem herkömmlichen Handel und e-commerce besteht darin, dass Geschäfte nicht wie üblich durch persönliche Anwesenheit und direkte Übergabe von Zahlungsmitteln gegen Ware abgewickelt werden, sondern sich Nachfrager und Anbieter in einem anonymen Rahmen gegenüberstehen. Somit kann die Vetrauenswürdigkeit von Anbieter und Nachfrager erst nach erfolgreicher Abwicklung des Geschäfts verifiziert werden.

Desweiteren muss zwischen business-to-business und business-to-consumer-Geschäften unterschieden werden. Bei ersterem handelt es sich um den elektronischen Handel zwischen Unternehmen (nachfragende Unternehmen sind hierbei z.B. Wiederverkäufer, Weiterverarbeiter) und bei letzterem um den Handel

[1] E-Commerce-Business in digitalen Medien, Gerhard Schreiber, Luchterhand 1998
[2] E-Commerce-Business in digitalen Medien, Gerhard Schreiber, Luchterhand 1998
[3] Wettbewerbsstrategien für e-commerce, Haertsch, 2000, 2

zwischen Unternehmen und Konsumenten (Endverbraucher). Im folgenden soll aber nur die B2C-Seite betrachtet werden.

Im Vorgriff auf im weiteren Verlauf verwendete Statistiken sei an dieser Stelle noch die Begriffsdefintion des Statistischen Bundesamtes genannt:

Demnach „beinhaltet e-commerce solche Transaktionen auf einem Markt,

- bei denen nicht nur das Angebot elektronisch offeriert, sondern auch die Bestellung bzw. die Inanspruchnahme elektronisch unter Verwendung eines computergestützten Netzwerks erfolgt, und
- durch die der Austausch von wirtschaftlichen Gütern gegen Entgelt (z.B. Kauf, Miete, Pacht) begründet wird."[4]

## 2.2 Was macht Online-Shopping attraktiv

### 2.2.1 für die Konsumenten

Der Hauptvorteil des e-commerce für den Konsumenten ist der convenience-Aspekt. Dieser äussert sich in der jederzeitigen Zugänglichkeit zu den Internetseiten der Anbieter, welche unabhängig von den Ladenöffnungszeiten ist und der Zugang von jedem beliebigen Ort aus erfolgen kann. Ausserdem entsteht eine Zeitersparniss durch die Vermeidung von Anfahrt und Parkplatzsuche und es wird weniger Zeit für den Kaufprozess selbst benötigt und Produkte können schneller verglichen werden. Bei einem Einkauf direkt vom Hersteller können ausserdem Kosten durch geringere Preise gespart werden und für eine Vielzahl der Kunden dient „Surfen" im Internet der Unterhaltung und erhöht somit den Freizeitspass. Viele Kunden bevorzugen den Kauf im Internet auch deshalb, weil hier eine grössere Produktvielfalt gegeben ist.

(Artikel von Christian Dach (www.ecc-handel.de) „InternetShopping – Besonderheiten bei Einkaufsstätten- und Produktwahl, Abb. 3)

### 2.2.2 für die Anbieter

[4] Wirtschaft und Statistik, 4/2003

Für die Anbieter ergeben sich aus der Teilnahme am e-commerce Möglichkeiten, Umsätze bzw. Marktanteile zu erhöhen. So entsteht durch die Nutzung des Internet eine neue Form der Kundenansprache und Kundenpflege, d.h. es können neue Kunden gewonnen werden und die bestehende Kundschaft kann besser betreut werden.
Vor allem kleinere Unternehmen können durch die Internetpräsenz eine grössere geographische bzw. marktmässige Reichweite erlangen und somit Marktanteile oder Umsätze erhöhen.
Für grössere Unternehmen hingegen spielt die Chance auf mehr Effizienz im Vertrieb eine grosse Rolle. So können zum Beispiel Werbeaktionen gezielt und wesentlich kostengünstiger durchgeführt werden, da Versand- und Druckkosten entfallen.
Heute ist es auch häufig eine Frage des Images, ob sich ein Unternehmen im Internet präsentiert. So spielt die Frage: „Was passiert, wenn wir nicht im Internet präsent sind“, eine wichtige Rolle.

### 2.3 Erfolgsfaktoren

Für den Erfolg im Internet gibt es eine Reihe verschiedener Faktoren, die beachtet werden müssen. Zu unterscheiden ist hierbei zwischen allgemeinen Faktoren, die vom Unternehmen selbst wenig beeinflussbar sind und den unternehmensspezifischen Erfolgsfaktoren wie angebotene Produkte, Art des Unternehmens oder Branche. Allerdings existieren noch keine allgemeingültigen Studien über unternehmensspezifische Erfolgsfaktoren und es kann deshalb nur aufgrund des Erfolgs einzelner Unternehmen auf mögliche erfolgsbeeinflussende Faktoren geschlossen werden.

#### 2.3.1 Allgemeine Erfolgsfaktoren

Um den Erfolg des e-commerce zu gewährleisten, müssen einige allgemeingültige Vorraussetzungen vorhanden sein.
So ist es für die Funktionalität des e-commerce zunächst wichtig, das die e-commerce Systeme effizient arbeiten. Darunter fallen folgende Kriterien:

- Potente und intelligente Suchmaschinen und Navigationssysteme:
  Erst durch die Einführung von leistungsfähigen und komplexen Suchmaschinen ist es den Anbietern bzw. Nachfragern möglich, eine Vielfalt von Waren zu präsentieren und diese auf effiziente Weise zusammenzubringen. Aufgrund der Informationsmengen im Internet spielen Suchmaschinen deshalb eine zentrale Rolle für den Erfolg des e-commerce und Wettbewerbspositionen werden von ihnen beeinflusst.

- Sichere und funktionelle Zahlungssysteme
  Um den reibungslosen Handel über das Internet zu ermöglichen, ist die Existenz von sicheren und funktionellen Zahlungssystemen unabdingbar. So wird z. B. in den USA vorwiegend mit Kreditkarte

gezahlt, was aber in Deutschland aufgrund der Angst vor Missbrauch der Kartendaten noch weniger verbreitet ist. Dieser Faktor erklärt den Vorsprung der USA im e-commerce gegenüber Deutschland, wo die Funktionalität der e-commerce-Systeme noch durch das Fehlen geeigneter Zahlungssysteme behindert wird.

- Guter Ausbaustand der Netzinfrastruktur
  Um die Hauptvorteile des e-commerce, also Interaktivität und die realitätsnahe Präsentation eines breiten Angebots zu gewährleisten ist die Verfügbarkeit von leistungsfähigen Übertragungsnetzen und ein hoher Ausstattungsgrad der Haushalte mit Computern und Internetanschluss unabdingbar.

Als weiterer allgemeiner Erfolgsfaktor sei die Akzeptanz beim Kunden genannt. Diese generiert sich vorwiegend aus folgenden Kriterien:

- Erhöhung des Sicherheitsempfindens
  Der für die Konsumenten wichtigste Aspekt für die Entscheidung, am e-commerce teilzunehmen ist die Frage über die Sicherheit der Anwendungen. Hierbei geht es einerseits um die Sicherheit in der Übermittlung und Verwendung persönlicher Daten, die z. B. für Werbezwecke missbraucht werden können, als auch um die Sicherheit von Zahlungssystemen.

- Geringe Kosten
  Ein weiteres Entscheidungskriterium für die Konsumenten sind die Kosten der Nutzung des e-commerce. Hierzu zählen Nutzungsgebühren für Internet bzw. Telefon, Versandkosten aber auch die Kosten für die mit dem Suchen und Entscheiden verbrachte Zeit.
  Die Nutzungs- bzw. Anschaffungskosten gibt es derzeit viele verschiedene Modelle, meistens muss sich der Kunde zwischen  und Flatrate entscheiden. Die Gebühren hierfür liegen derzeit zwischen .... Trotz dieser relativ geringen Nutzungspreise sind für viele deutsche Haushalte gerade diese ein Grund, auf einen Internetzugang zu verzichten und somit nicht am e-commerce teilnehmen zu können.

- Kundenservice
  Durch die weitgehend anonyme Abwicklung von Geschäften im Internet gewinnt der Kundenservice für die Konsumenten an Bedeutung. Da in den USA z.B. schon kundenorientierte Geschäftspraktiken bestehen und die Kunden an Service gewöhnt sind, ergeben sich für amerikanische Firmen erhebliche komparative Vorteile. In Deutschland hingegen bestehen diesbezüglich noch erhebliche Defizite. Diese reichen von der Notwendigkeit einer Organisationsform, durch die ein umfassender Kundenservice umgesetzt werden kann bis hin zur kaum vorhandenen „Servicementalität“, d.h. dem Ernstnehmen des Dienstleitstungsanspruchs der Kunden.

Um die Akteptanz beim Kunden zu erhöhen sollten auch die bereits genannten Faktoren, die den e-commerce für den Kunden attraktiv machen, wie z.B. die 24-Stunden-Verfügbarkeit unbedingt garantiert werden.

Als allgemeine Faktoren, die für die Anbieter die Entscheidungsgrundlage bilden, am e-commerce teilzunehmen können folgende genannt werden:

- Existenz bereits erfolgreicher Internet-Anbieter
  Die Entscheidung, am e-commerce teilzunehmen fällt einem Unternehmen leichter, wenn bereits gute Lösungen ihrer Branche existieren und diese auch erfolgreich sind. So rief z.B. der Erfolg von amazon eine Reihe neuer Buchanbieter auf den Plan.

- Investitionsbereitschaft
  Da die Umsätze über das Internet in vielen Bereichen derzeit noch keine grossen Zahlen erreicht haben und es auch erst wenige Beispiele gibt, in denen Gewinne erwirtschaftet wurden, ist es unabdingbar, dass das Unternehmen ohne Aussicht auf schnellen Gewinn investiert und die Fähigkeit besitzt, Durststrecken zu überstehen.

Weitere Aspekte für die Funktionalität des e-commerce sind politische Rahmenbedingungen wie Fragen der Besteuerung oder Subventionierung von Neueinsteigern und rechtliche Regelungen bezüglich des Vertagsabschlusses oder des Datenschutzes.

### 2.3.2 Unternehmensspezifische Erfolgsfaktoren

Bisher gibt es noch keine allgemeingültigen Aussagen über unternehmensspezifische Erfolgsfaktoren, aber anhand von Beispielen über bereits erfolgreiche Internetanbieter kann die Aussage getroffen werden, dass Unternehmen, die im e-commerce erfolgreich sein wollen, sich durch besondere Merkmale bezüglich ihres Produktsortiments bzw. ihrer Organisations- und Managemantansätze auszeichnen müssen. Diesbezüglich können einige Grundvorraussetzungen genannt werden:

- am Markt müssen Absorptionsmöglichkeiten für die angebotenen Produkte bestehen
  Gute Absatzchancen bestehen auch im e-commerce nur für Produkte, für die ein hoher Bedarf besteht. Deshalb haben Innovationen bzw. bestehende, stark nachgefragte Produkte größere Chancen auf Erfolg als andere.

- gut eingeführte Markennamen
  Die Existenz von Markennamen erleichtert den Verkauf über das Internet, da der Kunde bereits eine Beziehung zum Produkt und/oder Unternehmen aufgebaut hat und die positiven Erfahrungen auf den

Vertrieb des Produktes im e-commerce überträgt. Es können aber auch speziell für den Vertrieb über das Internet neue Markennamen generiert werden, die sich durch eine exzellente Bedienung auszeichnen und somit auch in andere Märkte eindringen können.

- niedrige Kosten

  Um den e-commerce erfolgreich zu machen, ist es wichtig, den Kundenvorteil des kostengünstigeren Einkaufs zu garantieren. Vorraussetzung hierfür ist die effektive Gestaltung der Abwicklung aller mit dem Online-Shop verbundenen Tätigkeitsbereiche und somit die Kostensenkung beim Unternehmen.

- Garantie einer schnellen und effektiven Bestellabwicklung und Lieferung, Kundenservice

  Für den Kunden von grosser Bedeutung ist es, nicht lange auf seine bestellte Ware warten zu müssen, das diese sofort vollständig geliefert wird und dabei nicht mit Mängeln behaftet ist. Um den nachteiligen Faktor der Anonymität weniger gewichtig zu machen, ist ausserdem ein umfassender Kundenservice unerlässlich, denn der für die Kunden primäre Unterschied zwischen den Anbietern im Netz neben dem Preis ist der Service. Ein umfassender Service erweckt im Kunden das Gefühl der individuellen und persönlichen Betreuung, weshalb beim Unternehmen ausreichende Kapazitäten für die Abwicklung des Kundenservice vorhanden sein sollten, um Kundenbeziehungen aufzubauen und zu pflegen. Wichtig ist hierbei vor allem die Qualität und die Schnelligkeit des Service.

  Dazu zählen auch Zusatzinformationen, die dem Kunden das Gefühl vermitteln, ernst genommen zu werden. Bei UPS und Federel Express z.B. wurde dies durch eine einfache Anwendung, nämlich die Statusabfrage über den aktuellen Standort ihres Paketes, verwirklicht. Dies erhöht weder die Geschwindigkeit, noch die Qualität der Lieferung, jedoch ist der Kunde zufriedener und für das Unternehmen ergibt sich ein kaum merkbarer Mehraufwand, da diese Daten ohnehin ständig gespeichert werden.

- Verwendung aktueller Technik

  Nur durch die Verwendung aktuellster Techniken kann das Unternehmen sein Angebot im Internet optimal präsentieren, die Geschäftsabwicklung reibungslos gestalten und auch Fragen der Sicherheit optimal lösen.

- Optimale Präsentation des Angebotes im Netz

  Um über das Internet Produkte abzusetzen, spielt die Web-Präsentation eine wesentliche Rolle. Um das Interesse der Kunden zu wecken, müssen die Seiten ansprechend und unterhaltend sein, um den Spass am Surfen zu erhöhen. Desweiteren muss der Kunde die Möglichkeit haben, schnell und direkt auf das übersichtlich dargestelllte Angebot zuzugreifen und nicht lange nach dem gewünschten Produkt suchen zu müssen. Dabei ist es auch wichtig, die gesamte Produktpalette auf der Website zu präsentieren . Es sollten Komplettlösungen angeboten werden, d.h. die Seite sollte die Möglichkeit zur Auswahl und Bestellung über das Netz, sowie die Abwicklung des Kundenservices und das Anbieten von

Zusatzleistungen wie z.B. Versicherungen, Garantien, Hilfestellungen enthalten. Anbieter, die ihre Ware zwar über das Netz anbieten, bei denen die Bestellung aber per Fax oder Telefon erfolgen muss, werden auf Dauer nicht erfolgreich sein.

Es müssen auf der Seite alle zum Kauf relevanten Informationen gegeben sein, wie Preise, Zahlungs- und Lieferbedingunen und ein einfaches Bestellformular.

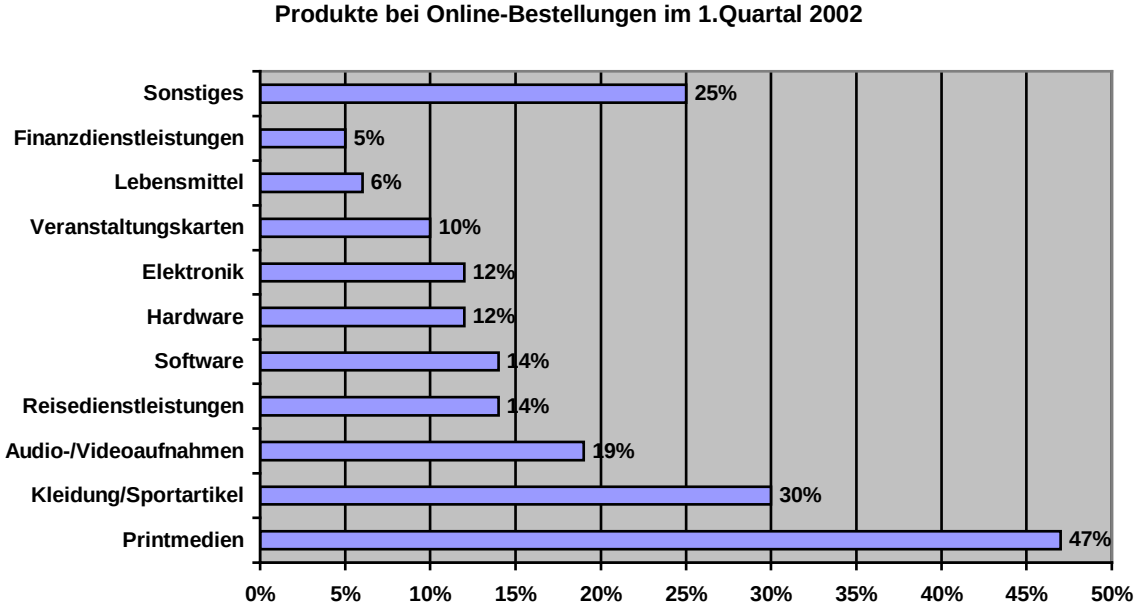
Ausserdem sind Bedienerfreundlichkeit und schnelle Ausführung der Transaktionen zu garantieren.

Es existiert eine Reihe weiterer Anforderungen an einen Online-Shop, die aber nicht Thema dieser Arbeit sind und auf die deshalb hier nicht näher eingegangen wird.

Über diese Erfolgsfaktoren hinaus gibt es noch eine Reihe weiterer spezifischer Kriterien, welche sich auf den Erfolg des online-Geschäftes auswirken. Diese lassen sich wie folgt gliedern:

#### 2.3.2.1 Art der Angebotenen Waren und Dienstleistungen

Aus einer Studie des Statistischen Bundesamtes geht hervor, dass die Online-Käufer hauptsächlich Artikel wie Bücher und Zeitschriften, Kleidung und Sportartikel, Musik- und Videoartikel sowie Reisedienstleistungen kaufen.




(Statistisches Bundesamt, Wirtschaft und Statistik 2/2003, S. 104)

Um daraus auf die Bedeutsamkeit einzelner Warengruppen für den e-commerce zu schließen, muss man zunächst eine Unterscheidung zwischen materiellen Gütern(Bücher, Kleidung, Lebensmittel) und digitalisierbaren Gütern(Software, Musik, Informationen, Nachrichten, Bilder, Videos) treffen. Innerhalb dieser Kategorisierung erscheint letztere Gruppe zunächst als geeigneter für den e-commerce, da die Auslieferung direkt über das Internet erfolgt und es für den Kunden nicht notwendig ist, diese unmittelbar zu erleben. Für erstere Gruppe gilt, dass die Kunden normalerweise einen Kauf im Laden vorziehen, da es bei

diesen Gütern wichtig ist, die Qualität, Frische, Grösse, etc. zu überprüfen und sich dies im Internet schwer realisieren lässt. Im folgenden werde ich näher auf die genannten einzelnen Bereiche eingehen.

Digitalisierbare Güter:

- Software

  Software ist aufgrund ihrer Beschaffenheit besonders für den Verkauf im Internet geeignet, da sie ohnehin schon in digitalisierter Form vorliegt. Zu beachten ist hierbei allerdings, dass Software kostenpflichtig ist, d.h. es sollten geeignete Zahlungssysteme mit der Website verbunden sein, aber es können z.B. kostenlose Updates angeboten werden. Der Vorteil des Verkaufs von Software besteht darin, dass die anzusprechende Kundengruppe „Computernutzer" meistens auch in die Gruppe der Internetnutzer fällt und diese gut mit einem Online-Shop erreichbar sind.

- Musik im Internet

  Der Erfolg von Musikanbietern wie napster, Music Store oder www.Mp3.com lässt die Schlussfolgerung zu, dass Musik ein vielversprechender Zweig des e-commerce ist. So wurden z.B. über den neu eröffneten Music-Store bereits in der ersten Woche mehr Songs heruntergeladen als für den ganzen Monat erhofft wurde und die enorme Nachfrage nach Musiktiteln bei dem neuen iTunes Music Store von Apple übertraf alle Erwartungen.

- Informationen und Nachrichten

  Für die erfolgreiche Vermarktung von Informationen und Nachrichten über das Internet seien vor allem Zeitungen wie die „Süddeutsche Zeitung" oder „Der Spiegel" genannt. Bei Informationen und Nachrichten handelt es sich um digitalisierbare Güter, deren Grenzkosten nahezu null sind und die auf einfachem Weg über das Internet vertrieben werden können. Über den neuen Absatzkanal Internet werden neue Kundengruppen erreichbar, die mit dem normalen Angebot bisher nicht angesprochen wurden. Das Internet ist also für Verlage ein optimaler zusätzlicher Vertriebsweg, dessen Zusatzkosten gering sind. Durch angebotene Zusatzleistungen wird die Attraktivität einer solchen Seite noch erhöht. Bei der Süddeutschen z.B. ist es für die Kunden möglich, ihre Abos interaktiv zu verwalten, Newsletter zu bestellen, an Diskussionsforen teilzunehmen und Antworten auf aktuelle Fragen zu finden.

Die im Online-Verkauf materieller Güter wichtigste Gruppe bilden Bücher, CD´s oder Videos. Während über die Hälfte der Deutschen einen Online-Kauf von Produkten dieser Gruppe in Betracht zieht, würden nur 16,1 % Lebensmittel kaufen. Die Ursache hierfür liegt darin, dass es für den Kunden bei Lebensmitteln essentiell ist, die Produkte unmittelbar auf Qualität und Frische zu prüfen. Erstaunlich ist allerdings, dass Textilien trotz des Bedürfnisses nach unmittelbarer Anprobe einen hohen Anteil am Online-Umsatz haben. Jedoch lässt sich dies durch die bereits gewichtige Position von Textilien im Versandhandel erklären. Bei Büchern, CD´s und Videos spielt das unmittelbare Erleben eine weitaus geringere Rolle, da sich diese Güter

in ihrer Aufmachung, Form und Qualität nicht von Anbieter zu Anbieter unterscheiden und weil alle wichtigen Produkteigenschaften auch medial vermittelt werden können (z.B. Leseprobe bei Büchern, Hörprobe bei CD´s).
An dieser Stelle soll auf die wichtigste für den Online-Handel relevanten materiellen Güter, nämlich Bücher, eingegangen werden.

- Bücher im Internet

  Im deutschen Internet-Buchhandel stehen derzeit die traditionellen Sortimentsbuchhändler neuen Konkurrenten wie Newcomern (branchenfremde Unternehmen oder Existenzgründer), Verlagen (Direktvertrieb) und vielleicht auch bald Autoren gegenüber, die ihre Bücher direkt an den Konsumenten verkaufen. Neben den bereits oben angeführten Gründen für den Erfolg des Online-Verkaufs von Büchern ist der, dass die Beratungsleistung von den Konsumenten als besser eingestuft wird als im stationären Handel. Sie haben ausserdem die Möglichkeit, auf ein umfassendes Angebot zuzugreifen, Leseproben, Rezensionen und Informationen zum Autor abzufragen.
  Derzeit erfolgreiche deutsche Internet-Buchhändler sind z.B. libri.de, buch.de oder kinderbücher.de, die aber alle drei nicht mit dem weitaus erfolgreichsten Anbieter, nämlich dem amerikanischen Online-Riese amazon mithalten können. Der schnelle Erfolg diesen Unternehmens liegt vor allem darin, dass es das erste seiner Sparte war und somit schon einen grossen Kundenstamm aufgebaut hatte, als sich die anderen noch überlegten, ob sich ein Internetauftritt lohnt. Amazon zeichnet sich vor allem durch eine einfache Navigation auf der Website, eine schnelle Reaktion auf Kündenwünsche und eine promte Lieferung aus. Um sich aber in Zeiten verstärkter Konkurrenz auch von Seiten deutschen Unternehmen weiterhin durchzusetzen, setzt amazon unter anderem auf eine Erweiterung der Produktpalette um Magazine und Zeitschriften und auf den Einsatz eines weiteren Tools auf der Internetseite, womit dem Kunden das Stöbern in Büchern, das Anschauen von Titelseiten, Buchrücken, Klappentext und verschiedenen Inhaltsseiten gestattet wird. Das bisherige Erstellen von Kundenprofilen aufgrund bisheriger Bestellungen und somit das speziell auf den Kunden abgestimmte Anbieten von best. Büchern wird nun über das Tool „Mein Shop“ abgewickelt. Hier findet der Kunde eine Produktauswahl, die sich an seinen bisherigen Käufen orientiert und der Kunde kann Produktlinien wählen, zu denen er künftig Informationen erhalten möchte.

#### 2.3.2.2 Versandhandel

Unabhängig von der Art der angebotenen Produkte haben Versandhandelsunternehmen einen komparativen Vorteil gegenüber dem konventionellen Handel. Sie verfügen bereits über die nötigen Organisationsformen um Bestellungen und Lieferungen effizient abzuwickeln und müssen um im e-commerce erfolgreich zu sein, nur noch eine angemessene Website generieren. Diese Unternehmen verfügen bereits über langjährigen Erfahrung im Versandhandel und darauf ausgerichtete, effiziente Liefermechanismen, sowie die für den

Kunden wichtige Effizienz der Abwicklung von Rücksendungen und Reklamationen. Desweiteren existiert bereits ein grosser Kundenstamm, dessen Vertrauen über Jahre hinweg gewonnen wurde. Herkömmliche Einzelhandelsgeschäfte hingegen müssen zunächst hohe Investitionen tätigen, um ihre Strukturen dem Online-Vertrieb anzupassen und müssen das Vertrauen der Kunden im Online-Verkauf erst durch optimale Abwicklung gewinnen.

Aktuelle Zahlen bestätigen den Erfolg des Versandhandels im e-commerce: Laut der Pressemitteilung des bvh konnten die Versender in 2002 gut 21,3 Milliarden Euro umsetzen, 661 Millionen mehr als noch in 2001. Die Gründe hierfür sieht der bvh neben der Eurotreue der Branche und den umfangreichen Serviceleistungen vor allem im e-commerce. So stieg der elektronisch erzielte Umsatz im Jahr 2002 auf 2,7 Milliarden Euro an, was einen Anteil von 13 Prozent am Gesamtumsatz ausmacht. Die Versender erwarten bis zum Jahr 2010 einen Anstieg diesen Anteils auf 20 Prozent am Gesamtumsatz[5]. Diese Prognose begründete bvh-Präsident Rolf Schäfer damit, dass die Versender über das Internet eine Vielzahl von Neukunden erreichen.

Auch für das e-commerce Schwergewicht KarstadtQuelle dient das Internet nicht nur als Bestellkatalog sondern auch der Gewinnung von Neukunden, die durch Kataloge nicht erreicht werden. So verzeichnete KarstadtQuelle einen Umsatzzuwachs im Versandhandel von 5,1 % auf 8,2 Milliarden Euro, wobei die Online-Nachfrage von 0,8 Milliarden Euro in 2001 auf 1,2 Milliarden in 2002 anstieg[6].

#### 2.3.2.3 Dienstleistungen

Obwohl es sich bei Dienstleistungen um digitalisierbare Güter (Banken, Versicherungen) bzw. medial vermittelbare Güter (Reisen) handelt, möchte ich auf diese an dieser Stelle gesondert eingehen.

Gründe für den Konsumenten, auf das Online-Angebot von Banken, Versicherungen und Reisedienstleistern zuzugreifen sind neben den bereits genannten Vorteilen digitalisierbarer Güter auch der convenience-Aspekt, d.h. die Möglichkeit sich bequem und schnell über vorhandene Angebote zu informieren, als auch Kostenaspekte. So sind Bankdienstleistungen im Online-Banking wesentlich günstiger als im herkömmlichen Vertrieb.

- Banken und Versicherungen

  Laut einer Studie von Forrester nutzen über 60 Millionen Europäer die Online-Angebote der Banken, was 18% aller europäischen Erwachsenen entspricht. Für den Bankkunden wichtig ist vor allem die persönliche und individuelle Gestaltung des Online-Private-Banking-Angebotes sowie die Benutzerfreundlichkeit und Schnelligkeit.

  Warum einige Banken jedoch erfolgreicher sind als andere, liegt wohl vor allem in der unterschiedlichen Kundenstruktur. Während Sparkassen bzw. Volks- und Raiffeisenbanken jeden Kunden annehmen müssen, ist es für Privatbanken nicht zwingend, einer Person ein Konto zu eröffnen. Somit besteht die

[5] bvh, Pressemitteilung v. 24.03.2003
[6] www.karstadtquelle.com/ir/73_2970.asp

Klientel der Privatbanken eher aus vermögenden Privatkunden, sowie Geschäftskunden. Einer Studie des Statistischen Bundesamtes zufolge, hängt die Internetnutzung und somit die Teilnahme am e-commerce auch von der Einkommensklasse und dem Bildungsstand ab. Ist es somit möglich, dass z.B. die Deutsche Bank und die Commerzbank aufgrund ihrer Kundenstruktur mehr Erfolg im Internetbanking haben? Leider liegen hierüber noch keine fundierten statistischen Daten vor. Allerdings konnte ich auf den Internetseiten dieser zweier Banken herausfinden, dass es Teil ihrer Strategie ist, die Möglichkeiten des Internet auszunutzen und auch hohe Investitionen in diesem Feld zu tätigen.

- Reisen

  Reisedienstleistungen eignen sich besonders für den Vertrieb im Internet, da auf diesem Wege die Angebote mit Hilfe zahlreicher Bilder und Links zu zusätzlichen Informationen über das gewünschte Reiseziel erweitert werden können. Die Darstellung des Angebotes in diesem Umfang ist im Reisebüro und anhand von Katalogen kaum möglich.

  Die Attraktivität des Buchens von Reisen über das Internet lässt sich anhand folgender Zahlen darstellen: Der Online-Umsatz der Reisebranche Westeuropas hat sich im Jahr 2002 um 55 % erhöht wobei 7,3 Milliarden Euro insgesamt umgesetzt wurden. Flugreisen hatten daran einen Anteil von 62 Prozent. Für die Online-Bucher ist es allerdings wichtig, auf der Website eines grossen, bekannten Reiseveranstalters oder bei einem bekannten Reisebüro zu buchen.

#### 2.3.2.4 Portale und Marktplätze

Eine weitere Möglichkeit eines Internetauftritts stellen Portale dar. Dies sind Webseiten, die dem Benutzer als Einstiegshilfe im Internet dienen sollen. Die geschieht anhand der vielfältigen Leistungen wie Suchwerkzeuge, thematisch geordnete Webadressen, aktuelle Informationen oder Kommunikationsdiensten. Ein weiterer Teil eines Portals können Elektronische Marktplätze sein, welche aber auch losgelöst von Portalen in erscheinung treten. Das Funktionieren eines Marktplatzes und somit seine Attraktivität hängt erheblich von seiner Nutzung ab, d.h. je mehr Nutzer, desto attraktiver. Um die ständige Inanspruchnahme zu fördern, bzw. die Kundenbindung zu erhöhen, bieten sich z. B. die Vergabe von Bonuspunkten für getätigte Transaktionen oder personifizierte Ansprachen in Form von Chats, Foren etc. an. Desweiteren erwarten die meisten Teilnehmer ein Mindestmass an Vertrauen, welches durch eine Registrierung erreicht werden soll.

Der wohl bekannteste Online-Marktplatz und das einzige profitable und stetig wachstende Unternehmen der Internetbranche ist derzeit eBay. Ebay bietet seinen Nutzern die Möglichkeit, auf seinen Webseiten Verkaufsangebote zu präsentieren und andererseits als Bieter für Angebote zu agieren. Für viele Konsumenten ist dies eine Möglichkeit, sich günstig Waren zu beschaffen oder für nicht mehr benötigte Gegenstände noch einen Preis zu erzielen. Der Erfolg von ebay liegt in der einzigartigen Geschäftsidee sowie in der automatischen Abwicklung fast aller Transaktionen ohne teure Überprüfung durch Personal.

Im Jahr 2000 wurde der deutsche Ableger des weltweiten Online-Marktplatzes eBay gegründet und ist bereits heute mit einem Handelsvolumen von mehr als einer Milliarde Dollar im 4.Quartal 2002 der Shopping-Favorit der deutschen Internetnutzer. Doch auch eBay arbeitet ständig an Verbesserungen des Angebotes. So wurde vor kurzem eine Partnerschaft mit dem Versteigerungshaus sothebys eingegangen, um den Absatz von Antiquitäten und Kunstwerken anzukurbeln. Eine weitere Auswirkung der erfolgreichen Geschäftsidee ist die, dass derzeit viele Einelhändler versuchen, ihre Angebote über eBay zu verkaufen. Dies bietet sich vor allem im Geschäft mit Restposten und Neuprodukten an. Die Händler zahlen dafür Gebühren in Höhe von 0,25 bis 2,40 Euro und eine Provision von 1,5 bis 4 Prozent an eBay. Nachteilig aus Sicht des Finanzamtes sind jedoch Umsatz- bzw. Einkommenssteuerausfälle, da die Händler durch Verkauf über eBay Steuern umgehen können, falls sie sich nicht als solche zu erkennen geben.

#### 2.3.2.5 Fachgeschäfte

Spezialgeschäfte können sich im herkömmlichen Handel selten durchsetzen und werden durch geringe Kundenzahlen oft zur Aufgabe gezwungen werden. Durch den Einsatz von e-commerce bietet sich für derartige Unternehmen eine grosse Chance, denn grundsätzlich besteht beim Kunden der Bedarf nach derartigen Geschäften, die jedoch häufig nur in Städten zu finden sind und ausserdem oft nicht den gehobenen Ansprüchen gerecht werden. Durch einen Online-Shop besteht für solch Geschäfte die Möglichkeit, einen weitaus größeren Kundenkreis anzusprechen, da durch das Netz eine geographisch größere Fläche abgedeckt werden kann und somit den Erfolg des Unternehmens zu sichern. Für die Kunden ergibt sich dadurch der Vorteil, dass die Qualität der Beratung nicht mehr allein von der Kompetenz eines Verkäufers abhängig ist, sondern der Zugriff auf fundierte und vollständige Informationen besteht. Ein weiterer Attraktivitätsfaktor besteht hierbei natürlich in niedrigeren Preisen, die sich von Seiten der Anbieter aufgrund grösserer Absatzzahlen verwirklichen lassen.
Beispiele hierfür sind Fachgeschäfte für Tierbedarf oder auch Sportartikelanbieter, so z.B. Bergsport, Reitbedarf, etc. .

## 2. Vertrauen im e-commerce

Die blosse Neuheit des e-commerce ist ein Faktor, der Misstrauen erweckt. So wirken sich schlechter Kundendienst, Vertrauensmissbrauch, Spamming und ähnliches in der Anfangsphase weitaus stärker auf das Bewusstsein der Konsumenten aus, als es bei herkömmlichen Handelsformen der Fall wäre.
Auch die Eigenschaft dieser neuen Handelsform, das Fehlen von physicher Präsenz wirkt sich negativ auf das Vertrauen aus. So weiss der Kunde oft nicht, um welche Art von Unternehmen es sich bei dem von ihm gewählten handelt, ob es ein Kleinhändler oder ein alteingesessenes, vertrauenswürdiges Geschäft ist. Auch die ständigen Veränderungen der Webpages, oder das plötzliche offline-gehen ganzer Websites verunsichert die Nutzer bezüglich der Erreichbarkeit der Anbieter im Falle von Produktproblemen und Streitfällen.

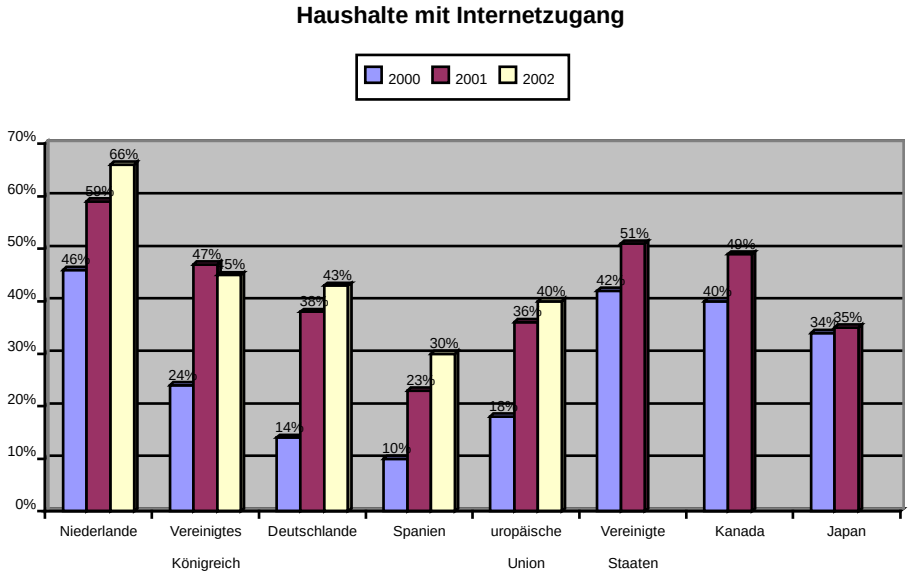
Ein weiterer Unterschied ist der Mangel an geeigneten Möglichkeiten zur rechtsgültigen Aufzeichnung von online-Transaktionen. Während im herkömmlichen Handel fast immer ein Kassenbeleg oder eine Quittung ausgehändigt wird, gibt es im online-Handel bisher noch keine passende Lösung.
Der eigentliche Vorteil des e-commerce, nämlich die Möglichkeit des Grenzüberschreitenden Handels wirft dabei auch Fragen bzgl. der Schlichtung von Streitfällen oder Problemen der Preisfindung, Lieferung, Besteuerung und der Währungsumrechnung auf.
Die Eigenschaft des Internet, dass den Verbrauchern Informationen normalerweise kostenlos zur Verfügung stehen, bringt für die bereitstellenden Institutionen jedoch das Problem der Finanzierung mit sich. Bisher wurde dieses z.B. über offene Werbung oder durch Erstellen von Benutzerprofilen für gezielte Werbung gelöst, was sich aber wieder negativ auf das Vertrauen der Verbraucher auswirkt.
Eine letzte Unsicherheit der Nutzer besteht in der Frage, inwieweit Informationen aus dem Netz zuverlässig sind. So kann der Kunde kaum einschätzen ob es sich bei den gefundenen Seiten um blosse Anfängerversuche oder ein professionell erarbeitetes Angebot handelt.
Eine erste Möglichkeit für die Anbieter, dass Vertrauen der Käufer zu erlangen, bietet die Zertifizierung des eigenen Online-Shops durch einen Dritten. Dabei handelt es sich um die Überprüfung eines Online-Shops hinsichtlich Leistungsmerkmalen und Sicherheitsstandards. In Deutschland vergibt das System Geprüfter Online-Shop des EHI solche Zertifikate. Die Händler werden hierbei jährlich auf die Einhaltung eines Verhaltenskodex überprüft. Auf europäischer Ebene ist ab sofort das Zertifizierungssystem Euro-Label im Einsatz.
Diese Maßnahme zusammen mit den zahlreichen neuen rechtlichen Regelungsen sollen das Vertrauen der Verbraucher stärken und somit den Erfolg des e-commerce steigern.

## 3. Probleme im e-commerce

### 4.1 Soziale Probleme

Die wichtigste Vorraussetzung um aus dem e-commerce einen effizienten und verantwortungsbewussten Markt zu machen, ist der Zugang der breiten Masse zum Internet.
Für die Konsumenten kann dieser von zu Hause aus (PC, Notebook, Handy mit Internetanschluss), am Arbeitsplatz, am Ort der Ausbildung bzw. über Internetanschlüsse öffentlicher Stellen (Schulen, Internet-Cafe´s, Bibliotheken) erfolgen. Um die Möglichkeit der Bevölkerung auf das Internet zuzugreifen darzustellen, sei an dieser Stelle eine Erhebung des Statistischen Bundesamtes über den Ausstattungsgrad privater Haushalte mit Gütern der Informations- und Kommunikationstechnologie gegeben. Demnach waren im Jahre 2002 rund 57 % aller deutschen Haushalte (ohne Haushalte von Selbständigen und Landwirten, bzw. einem mtl. Nettoeinkommen von 18000 EUR und mehr) mit PC´s ausgestattet und knapp 8 % mit Notebooks/Laptops (Ergebnisse der laufenden Wirtschaftsrechnungen der Jahre 1999 – 2002). Internet- oder Onlinedienste waren im Bundesdurchschnitt in rund 43 % oder gut 16 Mill. Haushalten vorhanden. Mit diesen Zahlen liegt Deutschland etwas über dem EU-Durchschnitt von 40 %, wobei aber andere Länder wie

z.B. Niederlande mit 66 %, die Vereinigten Staaten mit 51 % oder Kanada mit 49 % einen weitaus höheren Ausstattungsgrad mit Internet- oder Onlinediensten aufweisen[7].




Wichtig im Hinblick auf die Ausstattung der Haushalte mit einem Internetanschluss ist die Tatsache, dass dies lt. dem Statistischen Bundesamt auch von der Zusammensetzung des Haushaltes (mit oder ohne Kinder), der sozialen Stellung (Einkommenshöhe) und der persönlichen Situation bzw. Hauptbeschäftigung abhängt. So sind z. B. 100 % der Studierenden PC-Nutzer und davon 97 % Internetnutzer während von den Personen im Ruhestand nur 19 % einen PC nutzen bzw. 12 % das Internet.
Diese Zahlen lassen erkennen, dass durchaus nicht jeder Bevölkerungsteil die Möglichkeit hat, am e-commerce teilzunehmen und es daher wichtig ist, auf die Anforderungen dieser Personen weiterhin durch Angebote im herkömmlichen Handel einzugehen. Der hohe Anteil an Internetnutzern andererseits spricht dafür, den e-commerce auszubauen und Absatzchancen bzw. Kundennutzen zu erhöhen.

## 4.2 Technisch/Rechtliche Probleme

Derzeit existieren noch eine Reihe verschiedener rechtlicher Unklarheiten bezüglich des e-commerce, welche seine schnelle Entwicklung behindern. Die aktuellen Probleme beziehen sich dabei auf die Bereiche Rechtssicherheit der Geschäftsabwicklung, Datenschutz, Verbraucherschutz, Urheberrecht bzw. Steuerrecht.
Fragen der Rechtsicherheit bestehen bezüglich dem Vertragsabschluss im Internet und wie dieser rechtsgültig angebahnt und abgeschlossen werden kann. Regelungen hierzu finden sich vor allem im BGB und im Signaturgesetz.
Die Datenschutzprobleme ergeben sich aus der hohen Gefahr des Datenmissbrauchs durch Online-Anbieter, denn das Internet ermöglicht das sammeln von Kundendaten und das Abspeichern von Kundenprofilen.

[7] Wirtschaft und Statistik, 2/2003, S. 98

Deshalb sind Verbraucherschutzregelungen notwendig, um das Vertrauen der Konsumenten in den e-commerce zu steigern. Vor allem im Fernabsatzgesetz sind hierzu wichtige Regelungen zu finden.
Probleme des Urheberrechts ergeben sich aus der Eigenschaft des Internets, dass jede Person unentgeltlich Informationen verwenden kann und der Urheber dabei nicht die Möglichkeit hat, nachzuvollziehen, wer seine Informationen wie verwendet.
Schliesslich ergeben sich beim Steuerrecht Probleme im Bereich der Ertragsbesteuerung, wenn sich der Handel über nationale Grenzen hinweg erstreckt und für die Finanzämter ist die Sorge nach Umsatzsteuerausfällen ein aktuelles Problem.
Um diese Bereiche zu regeln, bestehen derzeit eine Reihe gesetzlicher Vorschriften.

### 4.2.1 Regelungen des BGB

Bei Vertragsabschlüssen innerhalb Deutschlands, d.h. Anbieter und Nachfrager sind Deutsche und es wird innerhalb von Deutschland geliefert, sind das BGB sowie die sonstigen weiter unten aufgeführten Gesetzestexte anzuwenden. Im andern Falle, ist zunächst zu klären, welche Rechtsordung anzuwenden ist.
Grundsätzlich gilt im deutschen Recht gleiches wie im herkömmlichen Handel: der Vertrag kommt durch zwei übereinstimmende Willenserklärungen, nämlich Angebot und Annahme zustande, wobei die Präsentation von Waren auf einer Internetseite noch kein Angebot des Verkäufers darstellt, sondern lediglich die Aufforderung zur Abgabe eines Angebotes. Als Angebot des Käufers gilt ein Mausklick oder eine e-mail, und die Annahme des Verkäufers kann in gleicher Form erfolgen. Allerdings sollte der Verkäufer die Vertragsannahme bei Verkauf an Privatpersonen aus Beweisgründen zusätzlich durch Fax oder Post zusenden. Der Kunde hat wie im normalen Handel auch, die Möglichkeit eines Widerrufs seiner Willenserklärung innerhalb von 2 Wochen.
Zusätzlich zu den bisherigen Vorschriften des BGB gilt das am 01.01.2002 in Kraft getretene **Schuldrechtmodernisierungsgesetz** (Teil des BGB), welches grundlegende Bestimmungen über den Vertragsabschluss im Internet enthält. Demnach ist es den Kunden technisch zu ermöglichen, Eingabefehler bei Bestellungen zu erkennen und zu berichtigen und der Zugang der Bestellung unverzüglich zu bestätigen, mit dem Hinweis auf die Möglichkeit der Einsichtnahme in die Vertragsbedingungen des Unternehmens sowie die AGB, falls solche Vertragsbestandteil sind, und deren Speicherort.
Einen weiteren neuen Bestandteil des BGB bildet das **Fernabsatzgesetz**. Dieses Gesetz beinhaltet wichtige Regelungen zugunsten des Verbraucherschutzes. Demnach bestehen folgende Unterrichtungspflichten gegenüber dem Verbaucher: Demnach muss der Unternehmer den Verbraucher über wichtige Vertragsbestandteile, wie Identität des Unternehmers, Anschrift, wesentliche Merkmale der Ware oder Dienstleistung, Art und Weise des Zustandekommens des Vertrages, Mindestlaufzeit des angestrebten Vertrages, Preis der Ware oder Dienstleistung einschließlich aller Steuern und sonstigen Preisbestandteile, Liefer- und Versandkosten und das Bestehen eines Widerrufs- oder Rückgaberechts informieren.Verletzt der Unternehmer diese Informationspflichten, kann der Verbraucher den Vertrag unbefristet widerrufen, da die

Widerrufsfrist erst dann beginnt, wenn der Unternehmer seine Informationspflichten erfüllt hat. Im Falle eines Widerrufs bzw. einer Rückgabe trägt der Anbieter die Kosten für Rücknahme und Rücksendung.

### 4.2.2 Elekronisches Geschäftsverkehr-Gesetz

Ein weiteres neues Rechtsinstrument ist das seit 2001/02 gültige Elektronische Geschäftsverkehr-Gesetz(EGG). Seine wesentlichen Inhalte bestehen im Herkunftsland-Prinzip und den Informationspflichten des Webseitenbetreibers.
Ersterer Teil regelt die rechtliche Zuordnung der Anbieter zu seinem Herkunftsland. So gelten für einen Webanbieter die Rechtsvorschriften seines Heimatlandes, also des Landes, in dem seine Niederlassung betreibt, wobei der Sitz des Servers keine Rolle spielt.
Der zweite Teil dieses Gesetzes beinhaltet weitere Regelungen zum Verbraucherschutz. Diese beinhalten Vorschriften wie die Pflicht zur Angabe schneller elektronischer Kontaktaufnahmemöglichkeiten, das Handelsregister sowie die Bezeichnungen der berufsrechtlichen Regelungen. Für weitere Transparenz der Angebote sollen ausserdem der Name und die Anschrift ´leicht erkennbar, unmittelbar erreichbar und ständig verfügbar´aus der Internetseite hervorgehen. Verstöße gegen diese Vorschriften können mit Geldbußen bis zu 50000 Euro geahndet werden.

### 4.2.3 Signaturgesetz

Um Verträge im elektronischen Handel abzuschliessen, bedarf es einem Verfahren das technisch sicher und rechtlich anerkannt ist. Diese Kriterien erfüllt die Digitale Signatur, die den Urheber von elektronischen Erklärungen erkennen lassen und die Integrität von Daten verlässlich darstellen soll. Die technischen Anforderungen an eine Digitale Signatur sind im Signaturgesetz geregelt:

§14 Technische Komponenten

(1) Für die Erzeugung und Speicherung von Signaturschlüsseln sowie die Erzeugung und Prüfung digitaler Signaturen sind technische Komponenten mit Sicherheitsvorkehrungen erforderlich, die Fälschungen digitaler Signaturen und Verfälschungen signierter Daten zuverlässig erkennbar machen und gegen unberechtigte Nutzung privater Signaturschlüssel schützen.
(2) Für die Darstellung zu signierender Daten sind technische Komponenten mit Sicherheitsvorkehrungen erforderlich, die die Erzeugung einer digitalen Signatur vorher eindeutig anzeigen und feststellen lassen, auf welche Daten sich die digitale Signatur bezieht. Für die Überprüfung signierter Daten sind technische Komponenten mit Sicherheitsvorkehrungen erforderlich, die feststellen lassen, ob die signierten Daten unverändert sind, auf welche Daten sich die digitale Signatur bezieht und welchem Signaturschlüssel-Inhaber die digitale Signatur zuzuordnen ist.

Weiterhin sind Zertifikatsverzeichnisse vor unbefugter Veränderung und unbefugtem Abruf zu schützen und technische Komponenten nach dem Stand der Technik hinreichend zu prüfen.[8]

Um mit einer digitalen Signatur Dokumente elektronisch zu signieren, benötigt der Anwender ein digitales Zertifikat, das ihn ausweist. Diese elektronischen Ausweise werden von sechs staatlich anerkannten Trust-Centern verwaltet. Jedes Zertifikat beinhaltet dabei einen öffentlichen und einen privaten Schlüssel, mit Hilfe derer die elektronische Signatur erstellt wird. Der private Schlüssel dient dabei der Erstellung einer digitalen Signatur, mit der Dokumente elektronisch versiegelt werden können um sie vor Fälschungen zu schützen oder an Kommunikationspartner zu übermitteln. Gespeichert wird dieser geheime Schlüssel auf einer Chipkarte oder er ist über eine PIN gesichert. Der öffentliche Schlüssel dient der Überprüfung der Echtheit und Herkunft des Dokumentes durch jedermann. Ausserdem kann der Absender durch signieren eines Dokumentes mit dem öffentlichen Schlüssel des Empfängers sicherstellen, dass nur dieser Zugriff auf dieses erhält.
Anwendung findet die digitale Signatur bereits bei dem Vertragsabschluss per e-mail. Laut Rechtssprechung kommen e-mail-Ausdrucken beim Abschluss eines anspruchsbegründenden Vertrages im Prozess kein Beweiswert zu, da e-mail-Dateien manipulierbar sind. Allein mit einer digital signierten e-mail-Bestellung liegt dem Internet-Händler demnach eine juristisch verbindliche Urkunde über den Geschäftsabschluss vor.

### 4.3 Zahlungsverfahren und Zahlungssysteme

Um eine weiteres Wachstum des e-commerce zu gewährleisten, ist es notwendig, das mangelnde Vertrauen der Konsumenten bei der Übertragung sensibler Daten zu erhöhen. Dabei ist nicht das Bestehen von Sicherheitslücken der Hauptgrund, sondern der mangelnde Kenntnisstand der Verbraucher.
Der Erfolg eines Online-Shops hängt demnach auch von der Wahl eines geeigneten Zahlungsverfahrens ab. Allerdings ist es nicht einfach, die Interessen der Anbieter und Nachfrager gleichzeitig mit einem Verfahren zu befriedigen. Der Hauptentscheidungsfaktor für die Anbieter, bei der Wahl eines Zahlungssystems ist hierbei die Minimierung von Kosten und Ausfallrisiken, während für die Kunden die Faktoren Sicherheit und Bequemlichkeit eine größere Rolle spielen. Ein weiterer Konflikt besteht in der Präferenz der Verkäufer für Vorauszahlung im Gegensatz zur Zahlung nach Erhalt auf Seiten der Konsumenten. Basierend auf den verschiedenen Anforderungen der Käufer bzw. Verkäufer ergibt sich folgende Rangliste der derzeit verbreitesten Zahlungsalternativen aus Sicht der Konsumenten:

1. Rechnung
2. Lastschrift
3. Nachnahme
4. Kreditkarte
5. Vorauskasse

[8] Gestetz zur digitalen Signatur, §14

Für die Anbieter ist die umgekehrte Richtung zutreffend.
Um die derzeit noch vorherrschenden traditionellen Verfahren für die Händler sicherer zu gestalten, d.h. die Gefahr von Zahlungsausfällen zu verringern, können Verfahren wie Plausibilitäts- und Adressprüfung bzw. Bonitätsprüfung/Scoring angewandt werden. Bei der Adressprüfung wird die Gefahr von Lieferungen an fiktive Adressen minimiert. Bei Bonitätsprüfung wenden sich die Verkäufer an Wirtschaftsauskunftsdienste wie Schufa oder Creditreform, um abzufragen, inwieweit Negativinformationen, wie z.B. laufende Mahnverfahren, Insolvenz, historisches Zahlungsverhalten, über den jeweiligen Geschäftspartner bestehen. Da diese Anfragen jedoch nicht ganz billig sind, bietet das Scoring eine Alternative. Hier werden Kunden in Risikoklassen eingestuft, die abhängig von Wohnumfeld, Wohnlage und der wahrscheinlichen Kreditwürdigkeit gebildet werden.
Für die Händler sicherere Zahlungsalternativen sind z.B. Zahlung per Telefonrechnung, Kreditkarte, Mobile Payment.
Die Zahlung per Telefonrechnung bietet sich vor allem für die Abrechnung digitaler Güter, die direkt über das Netz geliefert werden. Dies könnte z.B. durch das Verwenden kostenpflichtiger Telefonnummern umgesetzt werden.
Eine andere Alternative bietet das M-Payment. Bei diesem Verfahren gibt der Kunde zum Bezahlen seine Mobilfunknummer an, woraufhin er per Kurzmitteilung über den Rechnungsbetrag informiert wird. Diesen muss er nun noch mit einer PIN bestätigen. Die Vorteile dieses Verfahrens liegen in der Kundenbindung für die Mobilfunkunternehmen, in der Sicherheit durch die SIM-Karten und der Geheimzahl. Probleme der Umsetzung finden sich allerdings bei Prepaid-Kunden ohne Laufzeitvertrag.
Um dieses Zahlungsmodell durchzusetzen, haben Vodaphone, T-Mobile, Orange und Telefonica Moviles ein Gemeinschaftsunternehmen gegründet, das eine einheitliche Abrechnungsplattform für das Bezahlen per Mobiltelefon erarbeiten soll.
Eine neuere Entwicklung im Bereich der Abwicklung von Rechnungsprozessen stellt das „Elektronic Bill Presentment and Payment“(EBPP). Darunter wird die sichere, vollständig elektronische und integrierte Übermittlung, Präsentation und Bezahlung von Rechnungen verstanden. Im Unterschied zur herkömmlichen Rechnungsabwicklung erfolgt nunmehr die Übermittlung von Rechnungs- und Zahlungsdaten elektronisch, wobei die DV-Systeme der Transaktionspartner integriert sind. Auch die Übermittlung der zahlungsrelevanten Informationen an die Bank sowie die Information über den Zahlungseingang erfolgen elektronisch, wobei Abweichungen zwischen Rechnungs- und Auftragsdaten bzw. Rechnungs- und Zahlungsdaten automatisch vom System erkannt werden.
Um EBPP erfolgreich zu machen, ist die Verwendung international verbreiteter Standards sowie der konsequente Einsatz durch alle an der Wertschöpfungskette beteiligten Partner unumgänglich. Bisher scheiterte der Erfolg dieses neuen Konzepts häufig an der schwierigen Intregration des jeweiligen EBPP-Systems in die bestehende IT- und Organisationsstruktur des Unternehmens.

## 6. Zusammenfassung, Aktueller Stand und Ausblick

Zusammenfassend kann man sagen, dass der Erfolg des e-commerce erheblich vom Vertrauen der Verbraucher, der Art der Produkte sowie der Ausstattung der Haushalte mit Internetzugängen abhängt.
Derzeit werden über den Absatzkanal Internet noch keine sehr hohen Umsätze erzielt, so erreichte der Einzelhandel im Jahr 2002 8 Mrd. Euro Umsatz (1,6 % vom Gesamtumsatz) und der Versandhandel mehr als 10 % am Gesamtumsatz. Dies betrifft vor allem die Quelle AG mit einem Anteil von 17 % am Gesamtumsatz (608 Mill. Euro) und die Neckermann Versand AG erzielt derzeit über 10 % des Gesamtumsatzes über das Internet und hat damit die Profitgrenze im Jahr 2000 überschritten.[9]. Allerdings gab es bisher jedes Jahr Steigerungsraten in den Online-Umsätzen.
Es besteht also durchaus noch Potential im Bereich des Vertriebs über das Internet, welches aber vor allem von der Steigerung des Vertrauens der Verbraucher abhängt. So gaben bei einer Befragung des Statistischen Bundesamtes über die Gründe nicht am Online-Shopping teilzunehmen 44% der Befragten Sicherheitsbedenken bei Kreditkartenzahlungen und 37% Befürchtungen des Missbrauchs persönlicher Daten an. Weitere Zweifel bestehen bei 31% über die Durchsetzbarkeit von Reklamationen sowie bei 25% Bedenken bezüglich der Lieferbedingungen und des Rückgaberechts oder mangelnde Kenntnisse bei 23% [10].
Negativmeldungen wie z.B. die Schließung des Online-Supermarktes von Otto bestätigen die Vermutung, dass der Erfolg auch erheblich von der Art der Produkte abhängt. Der erst im Jahr 2002 neu eingeführte Lebensmittelversand Otto-Supermarkt.de musste aufgrund des ruinösen Preiswettbewerbes wieder schließen. Damit erging es dem Versandhaus ähnlich wie dem US-Lebensmittelversender Webvan, der bereits im Sommer 2001 Konkurs meldete. Auch Versender wie Karstadt, Kaufhof, Spar, Einkauf 24 oder Netkonsum haben diesen Bereich bereits wieder aufgegeben (www.ecc-handel.de). Somit ist der Bereich Lebensmittel für den e-commerce eher nicht geeignet.
Hinsichtlich der prognostizierten Entwicklung des e-commerce seien die Daten einer aktuellen Studie von eMarketer angeführt: auch für die nächsten Jahre wird ein anhaltendes Wachstum des europäischen e-commerce erwartet, dabei sprechen die Analysten von einem Anstieg des internationalen Handelsvolumens auf ca. 243 Milliarden US-Dollar (für 2003 ca. 60,7 Milliarden US-Dollar).(www.ecin.de/news/2003/07/04/05947/)
Der e-commerce ist noch nicht am Ende seiner Entwicklung, welche von anfänglichen Euphorien über zwischenzeitliche Tiefs wieder zu neuen Hoffnungen geht. Es ist erkennbar, dass es in diesem neuen Geschäftsfeld neue Erfahrungen zu machen gilt, um aus Fehlschlägen sowie Erfolgen auf optimale Strategien in diesem Bereich zu schliessen.

[9] Wirtschaft und Statistik, 4/2003, S.314/315
[10] Wirtschaft und Statistik, 2/2003, S.104

Literaturverzeichnis:


Bücher: Preißl, Haas: E-Commerce-Erfolgsfaktoren von Online-Shopping in den USA und in Deutschland, Duncker & Humblot

Gerhard Schreiber: E-Commerce-Business in digitalen Medien, Luchterhand, 1998

Jörg Krause: E-Commerce und Online-Marketing, Hauser, 1999

Zeitschriften: Wirtschaft und Statistik des Statistischen Bundesamtes, Ausgaben 2 und 4 /2003

Internetlinks: http://www.ecin.de/news/

http://www.ecc-handel.de

http://www.e-Recht24.de

weitere: „Vertrauen im E-Commerce“ von Stefek Zaba, Hewlett-Packard Laboratories, Bristol, Großbritannien

Signaturgesetz

Pressemitteilung bvh vom 24.3.2003

Artikel „Wer kauft was im Web?“ aus Handelsjournal (2001), Heft 8, S.22-23